CHARLES UNIVESITY IN PRAGUE

FACULTY OF SOCIAL SCIENCES

Institute of Communication Studies and Journalism
Department of Media Studies

Deep Learning as a Socially Constructed Technology

Master's thesis

Author: **Jindřich Libovický**

Study programme: **Media Studies**

Supervisors: **Mehmet A. Orhan, Ph.D.** and **Mgr. Jaroslav Švelch, Ph.D.**

Defense year: 2017

Declaration

1. I declare that I carried out this master thesis independently, and only with the cited sources, literature and other professional sources.
2. I hereby declare that my thesis has not been used to gain any other academic degree.
3. I fully agree to my work being used for study and scientific purposes.

In Prague, January 5, 2018 Jindřich Libovický

Bibliographical entry:

LIBOVICKÝ, J. *Deep Learning as a Socially Constructed Technology* Master's thesis, Charles University: Prague, 2016. Faculty of Social Sciences, Institute of Communication Studies and Journalism.

Thesis length: 240835 characters

Anotace:

Předkládaná práce se zabývá hlubokým strojovým učením a umělou inteligencí jako sociálně konstruovanou technologií. Místo obvyklého pohledu, který vysvětluje vznik hlubokého učení jako důsledek stavu technologické reality, používám Bijkerův teoretický rámec a vysvětluji vývoj prostřednictvím hodnot a zájmů relevantních sociálních skupiny (veřejnost, lidé se zvláštním zájmem o technologie, IT specialisté a výzkumníci v oblasti umělé inteligence). Pro každou z těchto skupin jsem vybral anglicky psaná online media, která se na danou skupinu zaměřují a analzyzoval jejich obsahy v letech 2012–2016. Analýza ukázala posun od vědeckých k technologickým tématům v článcích cílených na výzkuníky v oblasti umělé inteligence a širokou veřejnost, které považují hluboké učeníza přelomovu technologii. Články cílené na čtenáře se zvláštním zájmem o technologie se umělé inteligenci podobně věnují, ale není jí zde přikládán zvláštní status. Stejně jako v případě médií cílených na IT specialisty, považují hluboké učení za technologii jako každou jinou.

Abstract:

The presented thesis focuses on deep learning and artificial intelligence as a socially constructed technology. Unlike the traditional view which explains the emergence of the technology via the inner state of technological reality, I try to follow Bijker's theoretical framework of social construction of technology and explain the development via interests of relevant social groups (general public, technology fans, IT specialists and AI researchers) and values they attribute to the technology. For each of the groups I selected several English- language online media and analyzed their content between years 2012 and 2016. The analysis showed a shift from scientific to more technological topics in articles targeted on AI researchers and broad public. In these articles, deep learning is presented as a breakthrough technology. Articles targeted on technology fans cover the news about artificial intelligence in details, but they do not attribute any special status to the technology. Similarly to IT professionals, they consider deep learning to be a technology as any other.

Klíčová slova:

sociální konstrukce technologie, hluboké strojové učení, umělá inteligence

Keywords:

social construction of technology, deep learning, artificial intelligence

Institut komunikačních studií a žurnalistiky FSV UK
Teze MAGISTERSKÉ diplomové práce

TUTO ČÁST VYPLŇUJE STUDENT/KA:

Příjmení a jméno diplomantky/diplomanta: Jindřich Libovický	**Razítko podatelny:**
Imatrikulační ročník diplomantky/diplomanta: 2014	
E-mail diplomantky/diplomanta: jlibovicky@gmail.com	
Studijní obor/forma studia: Mediání studia (kombinované)	

Předpokládaný název práce v češtině:
Hluboké učení jako sociálně konstruovaná technologie

Předpokládaný název práce v angličtině:
Deep Learning a Socially Constructed Technology

Předpokládaný termín dokončení (semestr, akademický rok – vzor: *ZS 2012/2013*)
(diplomovou práci je možné odevzdat nejdříve po dvou semestrech od schválení tezí)
LS 2016/2017

Charakteristika tématu a jeho dosavadní zpracování (max. 1800 znaků)**:**
Deep learning (machine learning using very deep nerual networks) as a technology emerged after a series of breakthrough publications in 2012. Since then it enabled huge improvement in many artificial intelligence tasks including image search, machine translation, speech recognition or even playing the game of Go. Whereas the technology practitioners must feel overwhelmed of what this technology has enabled, the majority of the end users may have not noticed that something dramatically changed behind the always-the-same user interface. People involved in this technology progress and aware of this development (the researches themselves, software developers or IT businessmen) form global communities where the blogs and other on-line media are the primary mean of communication and and a space where the communities are constituted. Because the deep learning it is a very young technology, there have not been many attempts to systematically describe the history of the technology. If there are some, they focus on the technical and technological conditions that enabled the technology, rather than the social one. As far as we know, no one attempted to interpret the Deep Learning from the social constructivist perspective.

Předpokládaný cíl práce, případně formulace problému, výzkumné otázky nebo hypotézy (max. 1800 znaků)**:**
In the thesis, we would like to explore the social construction of the Deep Learning technology within the framework of the relevant social groups introduced by Wiebe Bijker**.** In particular, we would like to use the on-line information channels including news servers and blog publishing technology news to identify what the relevant social groups for this technology are. By performing both the qualitative and quantitative analysis of the web pages we will try to identify what are the values, interests and topics the relevant social groups associate with the technology (e.g., easier life, more wealth, conquering nature, risk of unemployment). Finally, we will try to suggest how these values influenced the technology development and its applications.

Předpokládaná struktura práce (rozdělení do jednotlivých kapitol a podkapitol se stručnou charakteristikou jejich obsahu)**:**

- Artifiicial Intelligence and Deep Learning
  - Symbolic and connectionist paradigm since 1950's
  - Situation before Deep Learning emerged
  - Deep Learning since 2012
- Non-constructionist Interpretations of Deep Learning (technological determinism, Kuhnian view of scientific revolutions, mathematical artifact)

- Social Construction of Technology by Bijker
- Identification of relevant social groups as on-line-media audience
- Quantitative analysis of the selected material
- Qualitative analysis of the selected material
- Conclusions and hypotheses about social construction of Deep Learning

Vymezení podkladového materiálu (např. titul periodika a analyzované období)**:**
We will work with the articles published on web servers and blogs on the relevant topics written in English. We will start with the well known news servers (guardian.co.uk, cnn.com) and technology blogs (gizmodo.com, techcrunch.com, thehackernews.com) and proceed by following the links the articles declare to be their sources. We will download the articles published between years 2012 and 2016. Based on the results of the automatic quantitative analysis, we will select a subset of the material for qualitative investigation.

Metody (techniky) zpracování materiálu:
Firstly, we would like to identify what the on-line information channels people use to get information and communicate about the technology. Then the quantitative analysis of this webs will follow. We will download the content of these webs between 2012—2016 and use computational text analytics methods (Latent Dirichlet Allocation, tf-idf keyword extraction) to get the time series of the articles topics and their concurrence (this will be presumable thousands of articles). Based on that, we will chose which articles will seem the most interesting for deeper reading (probably tens of them) and perform the qualitative discourse analysis of the to identify the values and interests.

Základní literatura (nejméně 5 nejdůležitějších titulů <u>k tématu a metodě</u> jeho zpracování; u všech titulů je nutné uvést stručnou anotaci na 2-5 řádků)**:**

Bijker, W. E. (1995). *Of bicycles, bakelites, and bulbs: toward a theory of sociotechnicalchange*. Cambridge, Massachusetts: MIT Press. ISBN 9780262023764.

A book of three case studies explaining the social construction of science on three seemingly paradox cases of technological innovation in the 19th and 20th century. All of the technologies could have technically emerge much earlier, however it was the social environment that caused the development to be much less straightforward

Bijker, W. E. (2010). "How is technology made? – That is the question!". *Cambridge Journal of Economics* (Oxford Journals) 34 (1): 63–76.

A journal paper that summarizes another fifteen years research on the social construction of science since the publication of the book introducing the term using the already mentioned case studies from 1995.

Given, L. M. (Ed.). (2008). *The Sage encyclopedia of qualitative research methods*. Sage Publications.

This book brings a complete comprehensive overview of qualitative research techniques for social sciences. We will use the parts concerning the discourse analysis.

Bishop, C. M. (2006). *Pattern recognition and machine learning.* Company New York,

This is a textbook on Machine Learning techniques from which the Deep Learning has developed. In this book, all machine learning techniques are understood via the Bayesian statics which was the main paradigm of that time. This book can be used for a comparison how even the basic terms became re-interpreted.

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, *35*(8), 1798-1828.

This is a journal paper that summarizes the Deep Learning techniques and results achieved in 2012 and the early months of 2013. It also proposes what the next development could be. The author of this paper is a lead researcher who later invented major improvements in contradiction with his own improvements.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, *521*(7553), 436-444.

A journal paper presenting the state of the art in deep learning for the non-technical (but still scientific) auidence.

Diplomové a disertační práce k tématu (seznam bakalářských, magisterských a doktorských prací, které byly k tématu obhájeny na UK, případně dalších oborově blízkých fakultách či vysokých školách za posledních pět let)

Theses focusing on history of artificial intelligence:

- Iva Hostičková: Vývoj paradigmat výzkumu umělé inteligence (master thesis, Faculty of Arts, Charles University, 2014)
- Petr Šudoma: Významné směry v umělé inteligenci (master thesis, Faculty of Arts, Charles University, 2013)

Theses focusing on social construction of technology:

- Karel Svačina: "To jsou nějaké divné Windowsy": Případová studie socio-technické změny na české škole (master thesis, Faculty of Social Studies, Masaryk University, 2010)
- Jaroslav Švelch: Osmibitové "poblouznění": Počátky kultury počítačových her v Československu (disertation, Faculty of Social Sciences, Charles University, 2013)

Datum / Podpis studenta/ky

……………………….

TUTO ČÁST VYPLŇUJE PEDAGOG/PEDAGOŽKA:

Doporučení k tématu, struktuře a technice zpracování materiálu:

Případné doporučení dalších titulů literatury předepsané ke zpracování tématu:

Potvrzuji, že výše uvedené teze jsem s jejich autorem/kou konzultoval(a) a že téma odpovídá mému oborovému zaměření a oblasti odborné práce, kterou na FSV UK vykonávám.

Souhlasím s tím, že budu vedoucí(m) této práce.

……………………….

Příjmení a jméno pedagožky/pedagoga **Datum / Podpis pedagožky/pedagoga**

TEZE JE NUTNO ODEVZDAT **VYTIŠTĚNÉ, PODEPSANÉ** A VE **DVOU** VYHOTOVENÍCH DO TERMÍNU UVEDENÉHO V HARMONOGRAMU PŘÍSLUŠNÉHO AKADEMICKÉHO ROKU, A TO PROSTŘEDNICTVÍM PODATELNY FSV UK. PŘIJATÉ TEZE JE NUTNÉ SI **VYZVEDNOUT** V SEKRETARIÁTU PŘÍSLUŠNÉ KATEDRY A **NECHAT VEVÁZAT** DO OBOU VÝTISKŮ DIPLOMOVÉ PRÁCE.

TEZE NA IKSŽ SCHVALUJE VEDOUCÍ PŘÍSLUŠNÉ KATEDRY.

Contents

Introduction

Machine learning is a computational way of solving problem, not by explicitly programming a computer to do so, but instead by presenting a learning algorithms with a set of example inputs and corresponding desired outputs. It is a research area as old as the computer science itself. [1] Although machine learning had many successful applications in the past, including automatic fraud detection in the financial sector, email spam detection or machine translation, it was a series of breakthrough conference papers around 2010 that entirely changed the field (Hinton, Osindero, & Teh, 2006; Glorot, Bordes, & Bengio, 2011; Hinton, Deng, et al., 2012) and made one of machine learning subfields—*deep learning*—a big new area of interest of both researchers and businesses.

Ten years after the results were first published, it has undergone many changes and has been deployed in many applications including machine translation (Wu et al., 2016), Internet search, speech recognition (Amodei et al., 2016), image search (Gong, Lazebnik, Gordo, & Perronnin, 2013) or in personalized advertisement recommendation (Simonite, 2017)—technologies used by hundreds of millions people every day. Emerging technologies utilizing deep learning (autonomous cars, artificial intelligence playing the game of Go) have received massive public attention recently (Silver & Hassabis, 2016; Sang-Hun, 2016).

Deep learning as an advanced technology has a big potential and maybe has already started to have social, economical and political impact. The newly gained ability to mine meaningful information from huge amount of data can bring changes to the distribution of economical and political power (e.g. creates new barriers for start-up companies not having any data, data leakages pose

[1] The first known machine-learning, the Perceptron (Rosenblatt, 1957, 1958), was published in 1957, just two years before the general definition of machine learning as "*the ability to learn without being explicitly programmed*" (Samuel, 1959).

bigger danger of misuse) and even create a new power dichotomy: those who have and those who do not have the data.

Deep learning on one hand makes information easily accessible,[2] not only by better techniques for mining information from data, but also for by advances in quality of machine translation or image recognition. On the other hand, it can become a threat for people's privacy. Without deep learning, the only meaningful commercial use a Facebook-scale image collection is displaying some generic advertisements while users watch the images. With deep-learning-based technologies, an owner of such a collection can recognize all users by face and probably most of the places where the images were taken and objects in the images (Metz, 2016). This not only allows better targeting of advertisements (photos are a perfect source of information about ways of spending free time), but also brings new privacy issues at the same time. Companies operating in the field of machine learning seem to feel some social responsibility in this area which lead to founding non-profit company *OpenAI* (Metz, 2015) (backed by Elon Musk, Amazon and other companies) and the *Partnership on AI* (Hern, 2016) by Google, Microsoft, Facebook and IBM whose goal is to promote ethical use of AI which is beneficial for the whole society.

Computer scientists tend to explain the success of deep learning as an almost inevitable consequence of the state of the technological reality (High, 2016; LeCun, Bengio, & Hinton, 2015). The breakthrough just came at the moment when computers began to be able to store and process big amounts of data and their computational power became strong enough—the technological prerequisites just got ready. The 'Related Work' sections of the research papers tend to narrate the development as a linear chain of inventions that eventually led to innovations presented in every research paper. The same narrative is then propagated through various communication channels to developers, business decision makers and eventually to the public.

The illusion of a linear development arises from taking into account only the inner scientific and technological context of the innovations. If we step on the ground of social constructivism, which views the most of human development and acquiring knowledge as process mediated by social interactions,

[2] After all, Google's declared mission statement is to "*to organize the world's information and make it universally accessible and useful*" (`https://www.google.com/about/company`)

the role of social interaction and values the social actors bring there cannot be neglected.

The constructivists' objection would certainly be that "[...] the problem is once [people] start to expect linearity, they bind themselves to the retrospective distortions that the linear description inevitably require" (Bijker, 1997, p. 7). The narrative makes a false impression of a causal chain of inventions and discoveries, one predecesed by the other. Informal conversations at research and development conferences go beyond the strict internal technological logic of scientific papers (LeCun, 2016), economical aspects are taken into account, however other social factors are generally overlooked.

Deep learning has its roots in academia and government-funded research. Unlike many other technological innovations, most of the development of deep learning has not been done in secretive research laboratories of big companies, but within the realm of creative co-operation of both universities and companies on open-source projects. A reason for that may be that it is no longer the ownership of the source code what gives the enterprises a competitive advantage, but rather the data they own and keep confidential. Even if someone got the whole source code of Google, it would take them years to build sufficient infrastructure to start a 'parallel Google' on the same scale, moreover without the collected user data the quality of their services would be much worse.

Fortunately for the presented thesis, heavy usage of on-line media left traces that could now be analyzed. The methodology that the Media Studies had developed for analysis of media content provides an excellent toolbox for studying deep learning from the perspective of social sciences. The depiction of artificial intelligence in different media—scientific papers, blog posts targeted on professional communities and articles from more conventional media form a mix that can tell us a lot about which values the society connects with the technology.

In this thesis, I attempt to capture the rise of deep learning technology from a perspective of various groups. I use on-line media and social media content to identify the relevant social groups and the values they associate with deep learning and artificial intelligence during the period from 2012 to 2016. The global nature of the open-source movement underlined by English being virtually the only language used makes all important data easy to analyze. Another fact making the research easier is that the on-line communities created centralized places I can use as starting points for seeking for other resources. *GitHub* is currently the world-leading platform for storing and discussing source code issues and create a common place where researches and engineers meet. Many of the open source software packages have their mailing lists. A community help server *StackOverflow* plays a similar role for software developers. All these on-line community hubs help me identify what are the relevant social groups and what the on-line media of the social groups are—which are personal or institutional blogs for the more expert communities (e.g., blogs on *LinkedIn* for more business and management oriented readers, *Google Research* blog oriented more on the engineering community), as well as more institutionalized ways of publishing (servers like *Wired.com*, technological sections of traditional news servers). By delving into the community discussions I can easily spot which blogs and servers publish content for which social groups for which deep learning may have some meaning.

As a theoretical background for my research, I use well-established Bijker's conceptual framework of the theory of Social Construction of Technology (SCOT) (Bijker, 1997). With its notions of 'relevant social groups' and 'interpretative flexibility', it provides a good theoretical explanation of how new technologies emerge. Although the theory can be justifiably criticized for a lack of explanative power for periods when the technologies are already established in the society (Latour, 1992), we can disregard these objections because deep learning is still far from reaching that stage of development.

The rest of the thesis is organized as follows: Chapter 1 introduces the reader in the areas of artificial intelligence and deep learning. Chapter 2 summarizes the constructivist perspectives on technology and suggest how this framework can apply to deep learning. Chapter 4 then describes methodology of the conducted empirical study. The last two chapters present results of the quantitative and qualitative research done on the collected data.

Chapter 1

Artificial Intelligence and Deep Learning

The term 'artificial intelligence' (AI) can have different meaning for different people. There is an ever-moving boundary between what is 'just a computation' and what should be considered already 'intelligent'. Recently, a new buzzword 'smart' obfuscated this distinction even more. Problems that have been considered part of AI as finding the shortest path from one point on a map to another, are now considered to be a standard part of discrete mathematics with nothing intelligent in it. Thirty years ago, asking a smartphone in natural language for a way to a different city, would be definitely considered to be AI. What in fact happens is that machine-learned speech recognition transcribes the query into plain text and some keyword matching rules extract that you want to navigate somewhere and where the target is. The rest is just normal computation in a big geographic database. It is definitely a smart approach, but I would hesitate to call it artificial intelligence.

Apart from that, there might be different connotations for AI in various fields. For science-fiction writers, AI may be represented by killer robots getting out of control. For game developers, it may mean manual scripting behavior of human-like game agents. For software engineers at Google, it can mean programming an intelligent assistant mining information from users e-mail into their calendar (which will be probably labeled as *smart* as well).

This chapter provides the reader with a brief introduction to what is currently called AI within the computer science and its history since the advent of digital computers in the 1950s. In the following section, I describe the evolution of deep learning techniques in the recent years and their basic principles from the usual non-constructivist almost linear-development perspective.

1.1 Where does AI come from

One of the most frequently used definitions of artificial intelligence is *intelligence exhibited by machines. In computer science, an ideal "intelligent" machine is a flexible rational agent that perceives its environment and takes actions that maximize its chance of success at some goal* (Russell & Norvig, 2002, p. 1). O'Regan (2016) goes with his definition even further claiming that "*the long-term goal of AI is to create a thinking machine that is intelligent, has consciousness, has the ability to learn, has a free will and is ethical.*" (O'Regan, 2016, p. 250)

In a broader sense, people talk about AI in case of a machine that is supposed to exhibit a behavior that would require non-trivial cognitive effort if performed by a human. Searle (1980) distinguishes between 'strong AI' which is really able to understand what is going on and find itself in various cognitive states and 'weak AI' which just simulates having real cognitive abilities. Nowadays, when people talk about AI as a technology, it is always the weak AI, strong AI still remains a hypothetical goal computer science. The contemporary AI contains many subfields: natural language processing, computer vision, planning, machine learning, etc. Many of them became established as independent fields with their journals and international associations.

In the early times of AI, researchers believed that a key to achieve a truly intelligent behavior is mastering tasks which are difficult to achieve for humans, e.g., finding the best path on a map or playing chess. Nevertheless, it appeared that although these tasks are tremendously difficult for humans, they are relatively easy for digital computers. On the other hand, tasks which are natural for humans like speech or object recognition still resist attempts to be fully mastered by computers.

From the early days of AI, we can distinguish two major paradigms— computational and connectionist (Russell & Norvig, 2002, pp. 13–16). These two paradigms coexisted thorough the whole history of AI.

The first one is so-called *computational* (sometimes called also *symbolic*) paradigm. It is based on the observation that human intelligent behavior often manifests as manipulation with symbols. Tasks like inferring facts from an already existing set of facts, doing mathematics, natural language processing—

all of them can be easily seen in terms of intelligent symbol manipulation, even object recognition or speech recognition can be seen as transforming noisy signal into an unambiguous symbolic representation. Even the famous *Turing tests* views intelligence as an ability to manipulate symbols (Turing, 1950). To pass this test, an AI needs to persuade people chatting with the machine in a text-only regime that it is a human being, not a machine. Another support for the computation paradigm comes from linguistics. Chomskian and other formal grammars (Sgall, Hajicová, Panevová, & Mey, 1986) well rooted in mathematical logics brought a suitable theoretical toolbox for natural language processing. Computational paradigm usually relies on discrete mathematical formalism and mathematical logics.

The connectionist view, on the other hand, gets its inspiration in biology. The ability to interpret and manipulate symbols is just a consequence of network of densely interconnected neural connections. What they really do is immensely different from symbol manipulation which only emerges from the almost chaotic behavior of a network. Except for already mentioned neural networks there are other techniques belonging to this paradigm like evolutionary computing relying on the principle of natural selection or swarm intelligence—every time the intelligent behavior emerges from coordination and development of simple units.

During the history of AI, the dominant paradigm swapped a few times. Unlike natural sciences where paradigm shift is usually slow, computer scientists tend to switch the mainstream relatively easily. A reason for that could be the policies of grant agencies which are always more likely to support an approach that seems to be most promising at the given time. However, during the history there were always some people believing more in other than the prevalent paradigm that were able to present results that persuaded the community to shift the paradigm again.

Neural networks as proponents of the connectionist paradigm received most attention in the early stages of AI (late 1950s), in the early 1990s and most recently after 2006. Here, I refer readers interested in the development of other techniques to overview by Russell and Norvig (2002, pp. 18–28) and continue with history of neural networks and deep learning only.

1.2 What is Deep Learning

Deep learning is a machine learning method whose models are composed of multiple processing layers which learn to represent data on increasing level of abstraction. Recently, it has shown a big success in tasks like image recognition, speech recognition, machine translation (and other natural language related tasks) and DNA sequence analysis (LeCun et al., 2015).

Nowadays, the deep models are usually used for so called supervised learning. In this setting, the model is presented with example inputs (e.g. speech recordings) and the desired outputs (e.g. the speech transcriptions) during the training. The training process tries to minimize the error on the training examples. The ability to learn a suitable numeric representation for a given task, together with superior performance is what makes the deep learning models so attractive both for researchers and software developers. The previously used machine methods required finding input representation manually which is usually impossible without big expertise both in the particular task's field and the field of machine learning.

I will illustrate the advantages of deep learning on an example from computational linguistics. Imagine, we want to machine-learn a program that will classify parts of speech a Czech coherent text. We have a corpus of millions sentences where the words are correctly classified[1]. The reason we might want to have such a program is that it can be useful for instance for keyword spotting or as a text preprocessing for syntactic analysis.

If we used a classical supervised machine-learning model, we would need to come up with some informative features that will allow the model to learn something about the words. The features usually need to be represented as sequences of zeros and ones. We would probably start with signs whether the word could be member of some closed-class part of speech (pronouns, prepositions, conjunctions, etc.). We could also distinguish whether they are ambiguous (e.g., "s" or "u" are always prepositions, whereas "se" can be either a pronoun or a preposition). For other word categories, the most important features are their suffixes and endings. We should therefore add a 0/1 sign for every ending that we can find in declination and conjugation paradigms

[1]For that purpose, we can use e.g., the Czech National Corpus (`https://ucnk.ff.cuni.cz/english`) collected at the Faculty of Arts of the Charles University

(e.g., "-a", "-ou"). Endings of words and endings of the surrounding words are probably the most important features for open-class words.

To design such system, we of course need an expert knowledge of the Czech language. With deep learning model, the input will be just characters the words consist of. The only thing we will need is the data and enough computational power, no specific knowledge of the language is required. The model finds on its own what are the important features of the words it should look at.

Deep learning originates in study of artificial neural networks that started in 1950s. Before 2010s, a neural network would be used as one of the classical supervised machine-learning models I mentioned in the example. The major breakthroughs that allowed models to learn everything from scratch are briefly described in the next section.

1.3 The Early Days of Neural Networks

Originally, neural networks implemented what the neuroscientists believed was a simplified model of a biological neuron. In this model, the neuron collects and weights the input signals on its dendrites and if the weighted sum of the neurons input exceeds a threshold, it fires a signal to the neuron's output, see Figure 1.2. This model has been called the perceptron (Rosenblatt, 1958). An important motivation for modeling a neuron was the opportunity to acquire some knowledge about real neurons via computational simulation. Despite its simplicity, the perceptron appeared to be a strong machine learning model which was capable to learn many things (e.g., character recognition that is shown in Figure 1.1). The principle perceptron learning still has many applications. The drawback of the perceptron is that it requires information rich and possibly mutually uncorrelated inputs. Researchers at that time believed that if they built a network of such perceptrons, the ones in the lower layers of the network will be able to learn the relevant features for the higher layers taking more complex decisions (Bishop, 2007, pp. 225–227). It took decades before this hypothesis was empirically verified.

After the memory capacity and computational power of computers allowed it, the researchers started to experiment with connecting artificial neurons into more complicated networks. For the sake of computational efficiency and simplicity, the networks did not contain any feedback loops which would have

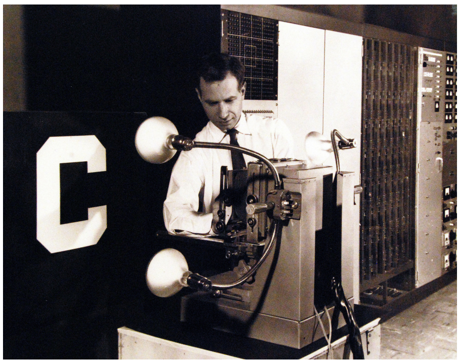


Figure 1.1: The percetron being used for printed character recognition. Photo from Bishop (2007, p. 196).

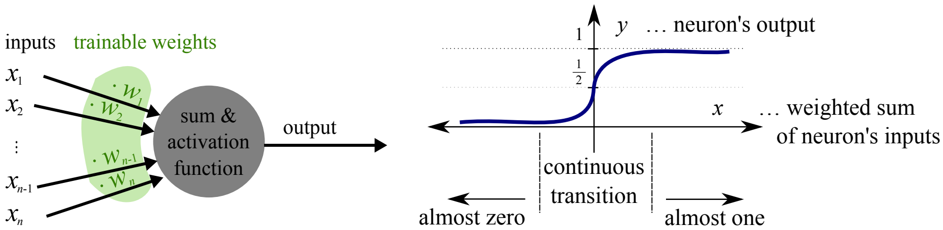


Figure 1.2: A scheme of a perceptron (left) and an example of an activation function (right). The weighted sum of inputs if on the x-axis, and the value fired by the neuron if on the y-axis.

made the computation numerically unstable. Unlike the biological neurons, the artificial ones receive their inputs in discrete time slices instead of operating in continuous time. Another traditional simplification is that the networks are organized into mutually interconnected layers which no connections within the layers, see Figure 1.3. When a network is organized in this way, computing neuron activation can be implemented efficiently using matrix multiplication. Moreover, a biological justification for this architecture was found in how the visual cortex is organized. (Fukushima & Miyake, 1982)

The other important moment in the development was invention of the back-propagation algorithm for training the neural networks (Rumelhart, Hinton, & Williams, 1988). Because of that, the neural network can be interpreted from as one big real-valued mathematical function. When the network is being trained, its inputs and desired outputs are fixed and we aim to minimize the error by changing the network parameters. Once the network is trained, parameters get locked and the model is treated as a function of its inputs. A strict separation between training and using is another point in which the artificial neural networks differ from the biological networks. This made the neural networks a

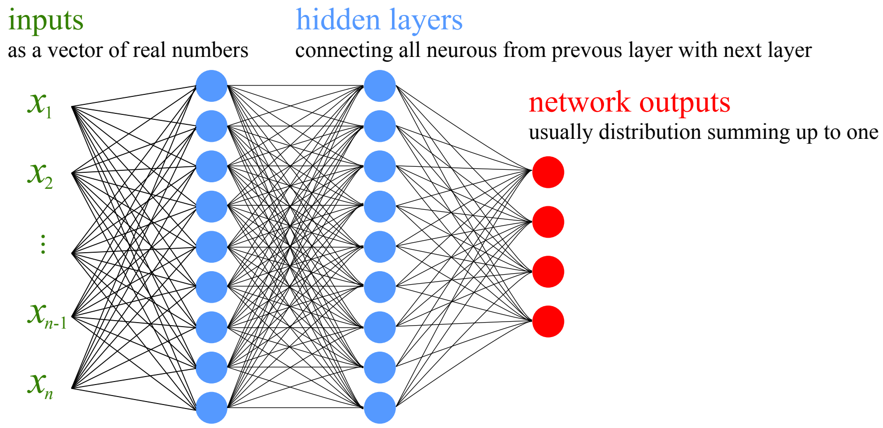


Figure 1.3: A scheme of a feed-forward neural network with the hidden layers.

practical and usable machine learning model, however it meant a resignation on the original goal to simulate biological neurons.

1.4 When Neural Networks Became Deep Learning

The previous section left the development of neural networks in the late 1990s or early 2000s. In this section, I will focus on the innovations that turned neural network research to what is now called deep learning. A remarkable thing about these breakthroughs is that many of these ideas have been tried unsuccessfully in the past. The availability of a bigger computational power enabled conducting much more experiments which eventually led to success.

The paper that returned neural networks to the center of scientific and later technological attention introduced layer-wise unsupervised pre-training of a model for handwritten digits' recognition task (Hinton et al., 2006). Each layer was independently optimized so that in a statistical sense it would best explain the numeric representation that appears in the previous layer. After preparing several layers this way, the model is trained using the standard back-propagation algorithm—in this case the process is called fine-tuning.

Although this is not the way the models are trained these days, it demonstrated that it is possible to train a deep network whose layers learn increasingly complex features of the input. Later, it was shown the model would learn the same task from random initialization, without the layer-wise pre-training, nevertheless it would take so much time that no one would let the computation finish having thought it simple did nothing.

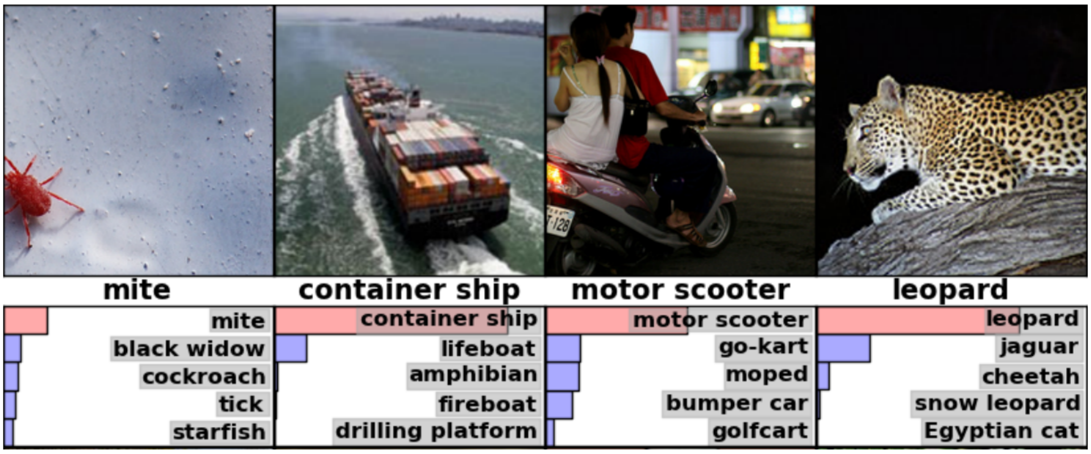


Figure 1.4: An example of object recognition from the AlexNet network (Krizhevsky et al., 2012)

Another impulse for further rapid development of deep learning came with the first success of deep convolutional network in the ImageNet challenge (Deng et al., 2009) (see Figure 1.4). The goal of the challenge is to recognize an object which is on a photograph, given approximately a million of photographs of one thousand categories used for training the model.

In 2012, the AlexNet network (Krizhevsky, Sutskever, & Hinton, 2012) beat the other methods by almost doubling the best performance from the previous year (with 62.5% first-best accuracy). Since then, the newer, deeper and by other clever tricks equipped network reached almost perfection on the challenge dataset. The 2016 best performing network ResNet (He, Zhang, Ren, & Sun, 2015) achieved accuracy of over 80.6 %.[2]

While using neural networks for natural language processing, the networks need to learn a numerical representation of the input words or characters. Although there is no direct supervision for these representations and the network learns to represent the words just in the way which is the best for the task it is supposed to do (e.g., machine translation or text summarization), the representation tends to have interesting properties. If we treat the representations as vectors in a multidimensional space, we might observe that they tend to form clusters according the words' semantic or grammatical properties (gender, case, etc.). Sometimes, they have even more interesting properties. A software tool called *word2vec* (Mikolov, Yih, & Zweig, 2013) by a Czech scientist, by

[2]The metric that is in fact reported in the challenge is so called 5-best error rate—the percentage of cases when the correct answer was not present in the top 5 guesses of the network. It was 15.7 % for AlexNet and 3.57 % for ResNet.

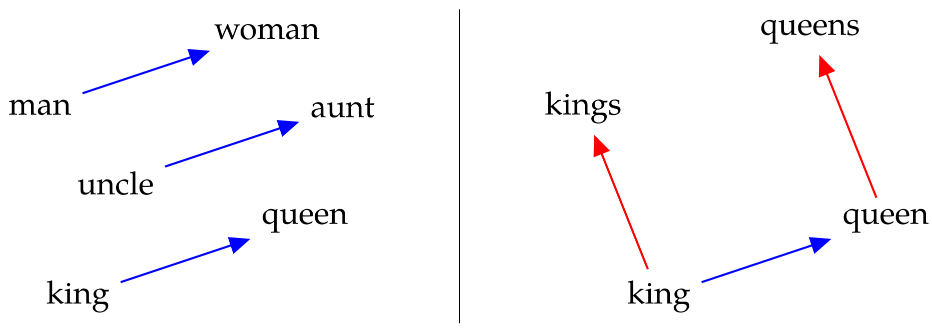


Figure 1.5: An illustration of geometric interpretation of vector arithmetics of word embeddings. The meaning shift can be often expressed as a constant shift in the vector space.

that time working at Microsoft Research, is able to infer word representation with semantic vector arithmetics. If we, for instance, subtract a vector for word 'man' from a vector for 'woman' and add it to a vector 'king', the closest word vector the result will be the vector for 'queen'. See Figure 1.5 for illustration.

Networks with feedback loops can be easily used for tasks like speech recognition. The input sound signal is split into discrete time steps and the network then estimates what phoneme corresponds to that particular piece of signal. A recurrent neural network [3]for speech recognition were able to decrease word error rate from approx. one percent (Hinton, Deng, et al., 2012).

One the most innovative ideas was using a recurrent network as a decoder (Sutskever, Vinyals, & Le, 2014). In such a model, a recurrent network is initialized with a vector encoding information about what should be generated. The recurrent network then in every time step outputs a symbol (typically a word or a letter), updates its internal state and puts its output as an input to the next step. Together with another important tricks (Bahdanau, Cho, & Bengio, 2014; Sennrich, Haddow, & Birch, 2016), this principle has successfully set up a new paradigm in machine translation while dramatically improving the performance (Wu et al., 2016).

Another astonishing property of deep learning is that it allows sharing data representations between vision models and natural language processing mod-

[3]Recurrent networks are well suited for processing of data of sequential nature as natural language. Unlike a feed-forward network, which processes an input through processing layers until it eventually provides an output and reseted every time they receive an input, the recurrent networks remember their state. Recurrent networks combine previous state statute? and a new input in every time step. After receiving an input the network updates its inner state and generates an output. For more details I refer reader to an introduction by Goodfellow, Bengio, and Courville (2016, pp. 363–383).

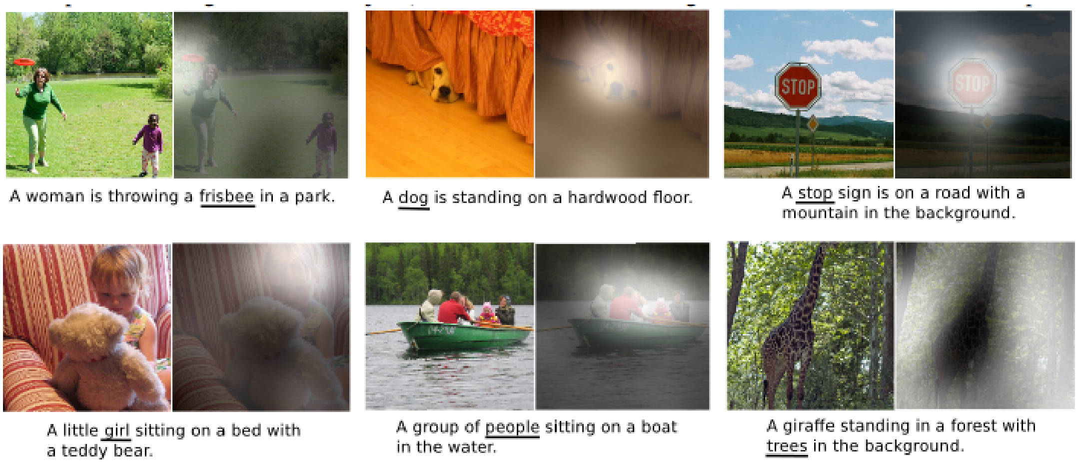


Figure 1.6: An example of generated image captions with output of the attention model for the underlined words. (Xu et al., 2015)

els which were considered to be fundamentally different research areas before. This allowed breakthroughs in tasks like image captioning (Vinyals, Toshev, Bengio, & Erhan, 2015; Xu et al., 2015), see Figure 1.6 for an example of the model outputs. The model is a simple combination of the convolutional network for image recognition on the input side and recurrent decoder originally designed for text generation in machine translation.

1.5 From Laboratory to Production

Companies try to move the deep learning empowered technologies to production as soon as possible. Google has recently launched deep-learning empowered machine translation (Wu et al., 2016). Facebook introduced a service telling visually impaired people what the content of the images may be (Metz, 2016). The advances in speech and object recognition and natural language understanding find many direct applications in mobile computing. Turning mobile phone into smart personal assistant is what all the major companies work on (Helft, 2016).

I can only speculate about reasons for which the companies invest in the products. Marketers probably believe that potential user will think such products will make their lives easier and more comfortable. Moreover, machines with cognitive abilities have been displayed many times in science fiction. Smart technologies coming with deep learning in some way match these science-fiction expectations, or at least these are the ideas the marketers might work with. Another reason might be the role of the personal assistant functionality. Having a personal assistant has been connected for a long time with having high social status. The deep learning-empowered technologies might be also seen as democratization of the privilege to have a personal assistant.

An interesting thing is that the founders of the field, which are among co-authors of most of the important publications, still understand the empirical success of the models as a kind of detour from the original aspiration to understand mind and intelligence via computational means (LeCun et al., 2015). This sounds even more paradoxical when we realize that most of the founders were hired by major technology companies: Geoffrey Hinton by Google (Hernandez, 2014b), Yann Le Cun by Facebook (Metz, 2013), and Andrew Ng originally by Google before he left for Chinese technological giant Baidu (Hernandez, 2014a).

The importance that technological companies attribute to AI was also apparent from one of Google's acquisitions when they bought an AI start-up DeepMind for 400 million British pounds after unsuccessful negotiations with Facebook (Gibbs, 2014). This newly acquired Google's division stood behind its AlphaGo software (Silver et al., 2016) based on deep learning which was able to beat the best Go player in the world (Silver & Hassabis, 2016; Sang-Hun, 2016).

Another factor that certainly helped a quick spread of deep learning in industry are other trends in information technologies which came around at the right time. Deep learning software is often hard to install and configure, and it requires a special hardware to run efficiently. Thanks to cloud computing, this is now a much smaller obstacle than it would have been only three or four years ago. It is easy to order a virtual machine in the cloud that has everything already pre-installed and set up. Moreover, well-defined remote call interfaces can keep the end-applications developers completely unaware of the inner "scientific computation" happening on a totally different machine in the cloud. It were also recent innovations in software and hardware engineering that allowed the deep learning technology to come to such a wide spread.

Chapter 2

Social Construction of Technology

In general, the *social construction of reality* (which is later applied on technological development) is a social theory that claims that humans grasp the world using the shared assumptions about the reality with the shared assumption standing for the reality itself. The basic theoretical background for understanding the constructivist viewpoint on technology is presented in Section 2.1.

The Social Construction of Technology (SCOT) is a well-established sociological theoretical framework applying these ideas to the technological world. While talking about technological innovations, it is easy to delve deeply in the technological nature of the innovations. Unlike the shared beliefs of decades or even centuries dead innovators, investors, consumers, etc., the technical details of the inventions are usually well-documented and if needed can be replicated at any time. While ignoring social context of the inventions, one can fall in a groundless surprise how the technology stunningly influenced the society, even though it may have been the social, political or economic environment that reinforced technological change itself. The Social Construction of Technology tries to provide a methodology to avoid such shortcuts in understanding the technological development. The theory is briefly summarized in Sections 2.2 and 2.3.

2.1 Theoretical Background on Social Construction of Reality

The *social construction of reality* is sociological theory that was introduced by a book of the same name by Berger and Luckmann (1966). The main question the theory attempts to answer is: how do people attribute meaning in their everyday life. In the reality of everyday life, people take most of the concepts, events and social situations for granted. Their meaning is not considered problematic at all. Practical use of everyday terms does not need to consider edge cases, does need to resolve inconsistencies if they do not collide with the practical use. This meaning can transcendent to the reality of theoretical thinking only with the difficulties, because theoretical thinking demands as unambiguous meaning as possible. Science often deals with this problem by postulating theoretical concepts which stand outside of everyday reality (e.g., mathematical concept of *set*), another way can be seen often in philosophy which problematize the everyday concepts beyond their everyday use.

When sociology wants to study social phenomena of everyday reality, it needs to conduct theoretical thinking over concepts of everyday reality which are not suitable for this purpose. For example, the notion of freedom as vaguely understood by most of the society members will be certainly different from a rigorous result of a thorough philosophical reflection of the notion of freedom. This is also the case of less abstract concepts like speed. Someone can easily say: "that is not speed what you are talking about because speed is in fact the derivative of a trajectory with respect to time". The theory of social construction reality overcomes this strange dichotomy meaning by replacing elusive meaning from everyday reality by the process of development, institutionalization of social phenomena, how they are known and how this knowledge is passed in the society. These can be rigorously described and thus provide a solid foundation for further theoretical work.

Objects in everyday reality (the phenomena about which we think are independent on our consciousness) have a meaning for us when they declare subjective intentions. For instance, a weapon always expresses a general intent to commit violence and does it in an intersubjective way—for everyone who knows how a weapon look like and what it can be used for. This is possible due to *typization* of its use. On top of objects which express subjective intentions directly, there are special objects which are called symbolic signs (and language as a sign system). A sign is something that carries subjective intention, which is typized but independent on immediate situation (it is possible to threat someone without showing an actual weapon, by just using words). The typization of language then also serves as a coercive measure to typize the experience itself.

Most of human activity usually get *habituated*, i.e., repeated activities eventually become patterns. The evident practical psychological advantage of habituation is that we do not need to re-invent everything when we do it and gradually get better in the activities. Mutual typization of habituated activities is called *institutionalization*. Social institutions emerge during the shared history, set normative patterns of human behavior and apply social control on it. The immediate practical advantage of institutionalization is that it makes it easier to predict what will other people do and thus make society more secure.

It can be illustrated on a (perhaps overly) simple example of a couple and their children. When a childless couple decides they will organize their everyday life in a particular way, their decisions are already a potential social institution. They know that they agreed on that and that they can change it any time they ant. For their children, however, all these decisions will become objective reality, something they were born to, similarly as if they were laws of nature. This objectivity (apparently something that is independent on our consciousness) is however a human creation which is paradoxically not perceived as human creation at all. The same process happens in much larger scale in the whole society.

Division of labour and other activities inevitably causes that different groups of people share different typizations of different activities and social institutions (with legal system or health care system being probably one of the most extreme examples of that). Mutually exclusive groups, especially when one is materially supported by the others, e.g., via redistributing of taxes within a state, need to legitimize their existence in a more general space of meanings shared by both groups. This particular aspect of social construction of reality is crucial in the next sections which discuss the meaning individual social groups attribute to technology.

Keeping together a society with many diverse groups living in different every day realities requires a value framework in which existence of the other groups is legitimate. Such ideological system, a *symbolic universes* is then a set of beliefs everyone is familiar with which and makes the institutionalized structure plausible and acceptable. Each society has mechanisms that keep everything in its right place and preserve the symbolic universe.

2.2 Wiebe Bijker and Constructivist Viewpoint on Technology

Thomas Kuhn starts his Structure of Scientific Revolutions (Kuhn, 1970) by the following thought: how it is possible that so many Aristotle's ideas are still valid nowadays, whereas his opinions on physics or biology seem to be naive or even stupid. These thoughts led him to inventing a notion of scientific paradigm and introducing his theory of scientific revolutions. Maybe these were similar thoughts that led Wiebe Bijker to develop his theory of social construction of technology. Having lived in the Netherlands, he must have been surrounded by bicycles and thought how was it possible that even though the technological means necessary for constructing a modern bicycle were available for merely half a century, it took until the beginning of the 20th century before the modern bike became a device usable as a means of everyday transportation. In his book *Of Bicycles, Bakelites, and Bulbs* (Bijker, 1997), it became a prototypical case of how social, political and economical aspects influence the technology as much as the technical aspects.

The Social Construction of Technology is a conceptual framework that started its development in the late 1970s as a critical reaction on so called technological determinism (Bijker, 1997). Technological determinism assumes that technology develops autonomously, having its own internal logic and it is therefore the technology influences to a great extend the social development. A famous example of this approach is work of Marshall McLuhan, who was able to explain in this manner many important moments of history of human communication and demonstrate how the technological innovation in human communication had fundamentally changed the way the western world worked (McLuhan, 1964).

The limitations of the deterministic view arise clearly when we start to ask questions like: was the Space Shuttle Challenger disaster in 1986 primarily a technical failure, an organizational mistake or a lack of funding (Bijker, 2010; Vaughan, 1997)? At the first sight, it may seem it was indeed a technological failure because it was the technology what failed. When this is analyzed more thoroughly, it becomes obvious that none of the suggested reasons is the primary one and that all the mentioned factors form a 'seamless web' (Bijker, 1997) of mutually interconnected relevant aspects. If NASA was organized differently, if there was a different atmosphere among the team, no reason to hurry, or different system of values shared among the employees, the technological failure, might not have happened at all.

Technological determinism applied in politics and policy making can have interesting and potentially undesirable consequences. Under such assumption, the only thing policy-makers can do is to follow how the independent world of technology is developing and if necessary, regulate the technologies that have emerged. Along with underestimating the role of political and economic environment, there is another political risks that can follow from technological determinism. The assumption that world of technology is constituted by expert knowledge of the technologists, firmly rooted in the objective world and therefore can be a source of "objective" policy-making advice (Bijker, 2006, p. 23–25), disregarding any implicit ideology driving the technological development.

SCOT applies the constructivist twist in sociology of technology. If we silently assume contemporary meaning and values attributed to technology together with universally valid objective engineering conceptualization of the technology development, the fact people did not invent a modern bicycle fifty years earlier is indeed understandable. In order to find out why that happened, we should focus less on the technological side of the problem and take into account values that were attributed to the technology during its development. They are of course vague defined and probably already forgotten in many cases. However, from the theory of social construction of reality, we already know they can be captured by the process of institutionalization of the technology use and other social phenomena associated with the technology.

2.3 Conceptual Framework of SCOT

From what was just said, it is clear that term technology in the conceptual framework of SCOT includes not only the technology artifacts and technological systems themselves, but also the knowledge about them and the practices connected with them.

The theory distinguishes four levels of analysis: a singular artifact, a technological system, a sociotechnological ensemble and technological culture. By trying to come up with a formal definition of these concepts (as in case of any other concepts), we would end up with complicated definitions with a long discussion on edge cases. For simplicity, I only illustrate the levels by describing typical phenomena that belong to the particular levels of analysis.

A singular artifact is a technological product, no matter whether a product invention (e.g., bicycle, smart phone) or a process invention (e.g., Bakelite, deep learning). The subject of research is *a story* of how a singular machine or a process is socially shaped. This is also the level of analysis that I focus on in the empirical study presented in the next chapters.

The key concepts the SCOT works with are the '*relevant social groups*' and '*interpretative flexibility*'. A relevant social group is any social group that *attributes a meaning* to the technology. Although, a membership in the groups can massively overlap, every relevant social group in fact sees actually a different artifact than the others. If we follow the example provided by Bijker (1997), for the high-wheeled bicycle that was common in the 1870s there were at least four relevant social groups:

- bicycle producers who wanted to make profit on the bicycle production,
- young athletic men who wanted to show off themselves,
- women bikers who faced gender discrimination because they could not wear proper clothes while riding a bicycle, and
- conservative moralist who has seen bicycles as an eccentric and dangerous things.

The existence of parallel interpretations is called '*interpretative flexibility*'. Notions of *relevant social groups* and *interpretative flexibility* are crucial for the empirical research introduced in the next chapters.

The later phases of the artifact development when it reaches a stable design are called '*stabilization*' and '*closure*'. At these phases, the artifact development can be narrated as a linear path of consecutive, mutually undermined inventions (or perhaps with few dead-end branches) leading to the artifact. Bijker (1997, p. 271) claims that after the closure, the history gets immediately rewritten. In the case of bicycles, this was reached by introducing the so called 'safety bicycle' whose design is stable since a century ago. This builds up a '*technological frame*.'

These terms can be seen as analogical Kuhn's conceptualization of scientific development. The 'technological frame' corresponds to Kuhn's 'normal science' and 'closure' to the moment when an 'alternative paradigm' becomes the dominant one.

Stabilization of the technological frame can be illustrated also on more recent examples. Few years ago, there were competing concept of touch screen smart phone and Blackberry smart phones with a physical QWERTY keyboard. Although at some point, the Blackberry phones were very popular. In the UK, they were so popular that they were attributed a crucial role in the 2011 London riots (Halliday, 2011). Having one these days would seem ridiculous.

The concept of '*technological system*' goes beyond the artifact and except the technical elements, it comprises the social, organizational, economic and political issues as well. When a technology grows, more a more capital, more technology and business effort are invested in its development—this builds up '*technological momentum*' which is the most important phenomenon on the system level. The bigger the momentum is, the harder is changing the course of the development. A good example of an unstoppable momentum is why we still use QWERTY keyboards despite the original technological motivation for this layout is no longer valid (Bijker, 2006, p. 29).

A '*sociotechnogical ensemble*' is a similar concept to a 'technological system', again considering a broader social context. The phenomena associated with the level of technology analysis are '*closed-in hardness*' and '*closed-out obduracy*'. The closed-out obduracy is a phenomenon experienced nowadays for instance by elderly people who do not have Internet access and involuntarily miss significant part of social and political issues. We talk about 'closed-in hardness' in cases when the technology is so rooted in the society people cannot think out of it and see any alternatives, e.g. when inhabitants of big American cities cannot really thing about commuting alternatives other than driving.

Chapter 3

Interpretation of Deep Learning

Although this thesis tries to present deep learning as a socially constructed technology, in this section I discuss other possible (and still important and relevant) views of how could the development of deep learning be reflected from a sociological or philosophical view. The first and probably the mainstream one can be categorized as a technologically deterministic view within social sciences. In contrast to it, I would like to put not only the SCOT perspective viewing deep learning as a technology, but also two other views in which deep learning could be interpreted as a mathematical artifact.

3.1 Technological Determinism

The story of deep learning as it is told in the big summary papers (Bengio, Courville, & Vincent, 2013; LeCun et al., 2015), and in a shortened form in introduction sections of all relevant research papers (and also Section 1.4 of this thesis) is basically a technologically deterministic story. All the important moments are explained via the inner technological or scientific logic of the field.

The impulse for the development of deep models came from the work utilizing the unsupervised pre-training of the models (Hinton et al., 2006). This particular technique used Bayesian statistics and graphical models which was one of the hottest approach in the AI community at that time. That probably helped to attract attention. Other crucial innovations followed. Changing the activation function from originally biologically motivated threshold functions to rectified liner units comes from the perspective of numerical mathematics (LeCun, Bottou, Orr, & Müller, 1998) became widely accepted. Dropout, making the training more robust by random disabling of neurons during training took inspiration from sampling techniques used also in Bayesian statistics (Hinton, Srivastava, Krizhevsky, Sutskever, & Salakhutdinov, 2012).

The last important piece of the puzzle is fast computation on graphics processing units (GPUs). GPUs were originally designed to accelerate graphics in computer games, where most of the image rendering is done by matrix multiplication. As mentioned earlier, if a neural network is organized into interconnected layers such that there are connections only between consecutive layers, exactly the same type of computation is needed as in computer games graphics. Krizhevsky et al. (2012) were able to gather all the innovations since 2006 and train a large network on two GPUs designed to accelerate computer games. It was the first time, neural networks proved their success on real-world photographs. All the methods before were tested on a dataset of handwritten digits of 14×14 pixels.

Telling the story in this way may create a false impression the neural networks were the first-class citizen of the AI research, not that they were mere outliers for a long time. An insight into this can be gained, if we look at extensive monographs summarizing the state of the art in machine learning in 2007 (Bishop, 2007, pp. 225–281) and 2013 (Russell & Norvig, 2002, pp. 727–737). The neural networks are indeed present in both books, but they view them from the perspective of Bayesian statistics which might seem obsolete today. The innovations that we now include in the straight story line of deep learning success are mentioned only marginally or not at all.

All steps leading to creation of the technology are determined by the previous state of the technological reality, until the technology became ready for deployment in the industry. First, step-by-step clever ideas, either borrowed from what was popular at that time in AI or determined by the hardware capabilities at that time, improved the networks' ability to learn, until the hardware was eventually ready for experiments in much larger scale than before.

At some point, a small parallel branch of research of recurrent networks merges into the story of deep learning. Recurrent networks are now used for processing sequential data, e.g., in speech recognition or machine translation. The most important idea in this branch was borrowed from electrical engineering and introduced logic gates into the recurrent connections (Hochreiter & Schmidhuber, 1997). The mathematical intuition and way of thinking of neural networks (developed in Munich) however does not fit well into the story from the previous paragraph (which happened in Montreal).

The unspoken need to find a linear story line in deep learning development lead to recent conflict of the former colleagues from CIFAR—Canadian Institute for Advanced Research (Geoffrey Hinton, Yann LeCun and Joshua Bengio) and professor Jürgen Schmidhuber from Technical University in Munich (now working at University of Lugano) calling the Canadian researchers the 'Deep Learning Conspiracy' and accusing them from plagiarism.

Hinton, LeCun and Bengio published a paper in 'Nature' (LeCun et al., 2015) summarizing the advent of deep learning. This emphasizes the importance of deep learning because computer science papers are only rarely published in 'Nature'. In his radical blog post (Schmidhuber, 2015), Schmidhuber claims that the main credit of Hinton and his colleagues is popularizing the deep learning, but not coming up with the main ideas. They re-invented many techniques that previously existed but not having direct technological application went unacknowledged and got forgotten or were rarely known outside specialized scientific communities. Scientists at CIFAR re-invented many of the algorithmic technique at a more suitable time with the promise of potentially monetizable impact in a relatively short time.

It is a conflict of two totally incompatible views. When the 'Deep Learning Conspiracy' wants to narrate the story of the deep learning development, they cite the works that have led them to the invention (and re-inventions) they have made. For them, these were the building blocks they used. The prior inventions which were made in a different theoretical, technological and social contexts simply cannot be part of their story. Not giving credit to original authors of the ideas may seem as plagiarism.

Schmidhuber's rigorous approach to searching the original authors of various ideas would be certainly more ethically appropriate. On the other hand, it would give credit to mathematicians who has nothing in common with deep learning and thus would tell a totally different story that never happened.

3.2 Deep Learning in SCOT

In this section, I show how the development of deep learning could fit the SCOT conceptual framework. I sketch a constructivist conceptualization of deep learning which is assumed through the rest of the thesis.

Deep learning is indeed not a material artifact, it is more a know-how which can be eventually objectified in a form of a source code. Nevertheless, in this stage of technology development, a running code or a software package is not something that could replace the know-how of the engineers. In spite of this, we can easily treat software based on deep learning (similarly to other process inventions and methodologies) the same way as material technological artifacts.

Research paper often highlight the ability of deep models to infer a good feature representation of the input (Bengio et al., 2013; LeCun et al., 2015). After all, from the research point of view, it is one of the most interesting properties of the models. While interpreted using linear algebra, the representations often have interesting properties. They often manifest something that seems semantic understanding of the model inputs which are often seemingly unrelated to the task the model is trained for (e.g., a machine translation models learn such a word representation that it forms clusters according to morphological categories of the words). This is certainly something the researches may value a lot and may be the reason to invest their effort in.

Representation learning is also of a big practical importance for *software development*: while using more traditional machine learning techniques, manual development of suitable quantitative characteristic, so called 'feature engineering', was the most tedious part of the development process. The ultimate goal in this sense would be a software that could be treated as a ready-made black box. It would only be presented some training data and do everything on its own. For developers with only a little interest in the learning techniques, this user-friendliness is certainly a value due to which they may consider using deep learning. The developers enter here as another relevant social group.

Once the models have been released in publicly available applications, *end users* became another relevant social group. This can be illustrated on a case when automatic image categorization by Google wrongly categorized an image of an Afro-American couple as an image of gorillas (Curtis, 2016). I was never an issue before for the developers, because they were only interested in the quantitative accuracy. Also, knowing the background they would have never attributed an algorithm such qualities as being racist. This makes an excellent example of different values seen in a technology. Legal departments of technological companies thus may see as an important property of the models, their stochastic and thus partially unpredictable behavior which can be easily overlook by the developers.

It is hard to say, whether deep learning achieved a 'stabilization' and a 'closure'. From the technical point of view it may seem so. However, deployment of the models will affect more groups whose interests and values will become important for the technology and this may form the technology further.

A 'technological framework' has developed as well. Originally, experimental software for deep learning experiments reached industrial quality, were publicly released and their APIs stabilized (Bergstra et al., 2010; Tokui, Oono, Hido, & Clayton, 2015; Abadi et al., 2016). Currently, there are several competing implementation of the deep learning frameworks. This can be interpreted as reaching some stabilization, but probably not as a closure because many groups have not said their last words.

Hardware producers offer specialized hardware for training deep learning models based on GPU technology originally designed for computer games (Gupta, 2014). Recently, Google announced development of a special hardware computation units for deep learning to work which will be less power demanding than the GPUs (Jouppi, 2016). The hardware infrastructure together with the stabilized software releases form a 'technological frame' of deep learning.

The growing momentum from the social perspective will be best seen in the result of the next chapter—increasing frequency of mentioning the technology outside the technical communities and connecting it with commercial products is the strongest evidence of that. The momentum can also soon reflect in users expectation: almost flawless speech recognition, handwriting recognition, huge improvement in machine translation can make users expect human-level performance in other cognitive tasks which would lead to further investments in the technology.

There is also a merely technological side of the momentum. Building the infrastructure for deep learning may cause that problems that would have been solved using other techniques will be now approached using deep learning despite not being the easiest solution, but the infrastructure will be ready.

3.3 Kuhnian Scientific Revolution

Deep learning technology developed from study of artificial neural network whose original goals was not to give a birth o a technology, but rather to help the neuroscience and philosophy of mind to find new pieces of knowledge via computational means. Most of the development of deep learning did not happen in industrial laboratories whose success is measured by the number of successful patent applications, but in a rather open academic environment. It justifies viewing deep learning more as an academic research field than a technology and it may be tempting to apply his view on scientific development on deep learning as well.

Kuhn's theory (Kuhn, 1970) tries to explain the scientific development via its underlying social dynamics in the scientific community. It assumes that when a new paradigm appears—a *scientific revolution* is going on—the clique of old-school scientists will never accept the new paradigm and the new paradigm cannot succeed before the proponents of the previous one eventually die out. Periods between two scientific revolutions are called the *normal science*. During that period researchers gather new observations and try to explain them using the current paradigm.

This is not entirely the case of machine learning, which is a very empirically driven field. Everyone in the field accepts that a class of models that systematically shows a superior performance over different classes of models is a way to go. The history of machine learning is full of sudden paradigm twists based on which class of models seemed to be currently the most promising. There are always researchers who either do not lose their trust or funding binds them to continue with a different paradigm which may or may not become dominant later.

The other reason why Kuhn's conceptual framework cannot be directly applied, may be the closeness of computer science and pure mathematics. In case of mathematics, Kuhn claims that it can be easy to change the paradigm if the new paradigm fits nicely into the aesthetic frame of the contemporary mathematics—simply if the other mathematicians consider a particular view on a problem nice or not (Kuhn, 1970, p. 155). Nevertheless, the rest of the conceptual framework can be applied quite accurately—we can observe paradigms changing, periods of revolutions and periods of normal science.

Most of the building blocks of the current deep learning come from the times when no one could anticipate the success of deep learning. Convolutional networks in form they are used were introduced in 1998 (LeCun, Bottou, Bengio, & Haffner, 1998), acceleration of learning by using different rectified linear activation is also from 1998 (LeCun, Bottou, Orr, & Müller, 1998), gated LSTM networks are even one year older (Hochreiter & Schmidhuber, 1997).

Usually, in order to get published, papers needs to pretend it is a part of normal science. Therefore, the papers (Hinton et al., 2006; Hinton, Srivastava, et al., 2012) that stood at the beginning of deep learning revolution interpreted the methods within the paradigm of Bayesian modeling. Their authors must have anticipated a revolution might be about the come, still did not indicate much in the papers. The unspoken rules of the community require the papers to be strictly technical and explain the place of the work within the state of the art, which actually means: do the normal science and explain how the work fits into the dominant paradigm. It was the strength of empirical evidence provided by Krizhevsky et al. (2012) that gave the authors ammunition to speak outside of the current paradigm.

All the crucial papers were published before the paradigm got established. Once it was there, these became a standard toolbox of the of normal science. Many innovations come from re-discovering or re-applying principles from the previous paradigms. Other innovations can be seen as systematic re-application of success stories well-established within the paradigm. For instance, so called highway networks (Srivastava, Greff, & Schmidhuber, 2015) re-apply the gating principle of gated recurrent networks to feed-forward networks or residual networks (He et al., 2015) use the principle of dropout mechanism which operates on level of singular neurons, on the whole layers of the network.

Terms of paradigmatic shift (which happened less than five years ago) and normal science (which is what we experience right now) thus can provide a good explanation of how innovation are made, but does not tell us much about how deep learning lives outside academia and how it works as a technology.

3.4 Kvasz and Patterns of Change

When we view deep learning as a set of mathematical tools rather than a technology, we may notice there are other regularities in its development common with the development of mathematics. Ladislav Kvasz (2008) introduced in his book *Patterns of Change* an unorthodox view on the development of mathematics that attempts to explain the Kuhnian social dynamics of the field not within social constructivism, but by searching for underlying phenomena of linguistic nature. According to Kvasz's views, the paradigm shifts are not driven by the social dynamics of the scientific community, but rather by the language of mathematics, its explanatory power and limitations. The innovations come from the language use and dealing with its limitations in scientific life. The social dynamics is interesting but inessential side effect. The resistance of old-school cliques is not driven by the fear of losing their respectable position, but just by a reluctance to accept new jargon, similar to elderly people refusing to accept the slang of the youth.

In Kvasz's theory, paradigmatic shifts or changes of a language frame are conceptual changes that alter the content of the terms the mathematicians work with. This allows explaining previously unsolvable problems in a stunningly easy way. On the other hand, it also allows inventing new mathematical objects which are impossible to grasp with the current conceptualization of mathematics. This is one of the patterns that appears repetitively thorough the history of mathematics.

An easily understandable example can be the situation in mathematics before Descartes invented the coordinate system. Mathematicians knew many tricks how to solve some types of algebraic equations. From time to time, someone invented another trick helping with another type of equations but a general framework was missing. They did not have a clue that the polynomial terms on one side of the equations can be interpreted as curves in a coordinate systems and that the roots are points where the curve crosses the x-axes (see Figure 3.1). In the geometric world, x^2 must be in square units, x^3 in cubic units—it certainly requires a lot of courage to deny some of the most fundamental rules of the Euclidean geometry and imagine a square as a curve in a coordinate system.

In Cartesian mathematics, on the other hand, we are able to generate trigonometric functions, which were unthinkable without a coordinate system. Proper explanation of these objects was nonetheless impossible within Cartesian geometry and mathematicians had to wait until Cauchy and Gauss formalized differential calculus almost two centuries later (Kvasz, 2008, p. 38–39).

$$
\begin{aligned}
f(x) = 10x^2 - x - 60 &= 0 \\
g(x) = 0.2x^3 - 2x^2 - 10x + 4 &= 0 \\
h(x) = -2x^2 - 10 &= 0
\end{aligned}
$$

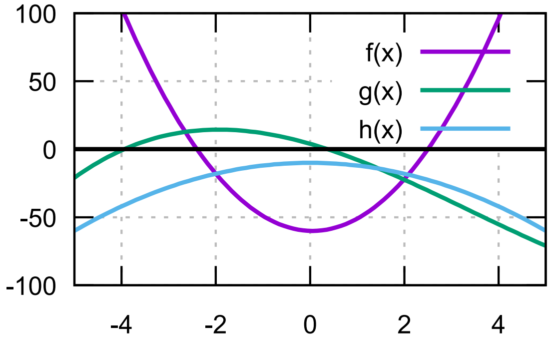


Figure 3.1: Examples of polynomial equations and associated curves.

The patterns Kvasz recognized in the history of mathematics appear in the development of deep learning as well. Currently, researchers lack formal mathematical understanding (they have indeed some intuition based on extensive knowledge of linear algebra, multidimensional calculus, numerical mathematics, graph theory and parts of high-level mathematics) of what particular parts of neural networks do. Unlike mathematics that requires rigorous proof of every new statement, researchers in deep learning often do not have means of proving new ideas right or wrong other than conducting an experiment on data. Reverse-engineering of the results might give them at least some insight. It works the same way as when medieval mathematicians had in mind particular way of solving an algebraic equation, they needed to try it out and see whether it works or not.

This does not mean that the current deep learning practitioners do not understand what they are doing or are not innovative enough to come with a holistic theory of machine learning. They just cannot think out of the current linguistic frame (or paradigm in Kuhn's words) because they lack the tools to express such theory. Deep learning model are studied empirically as if they were not mathematical artifacts but rather natural objects.

At the same time, we witness many attempts to grasp the models mathematically in many creative ways that often led to interesting innovation as if they were some tendency to re-code deep learning in order to understand it. Batch normalization (Ioffe & Szegedy, 2015) and residual networks (He et al., 2015) require a little re-coding, probably connected with how the networks are implemented in current object-oriented programming languages. These innovations view the models not as a network of individual neurons, but as a network of layers. The gating mechanism in recurrent networks (Hochreiter

& Schmidhuber, 1997) was mathematically justified using a graph-theoretical metaphor of creating free path for information during the network learning. The attention mechanism for direct probabilistic accessing the previous network states (Graves, Wayne, & Danihelka, 2014; Bahdanau et al., 2014) view the model as a machine that performs some tasks in a somehow mechanistic manner. The model is a machine, whose instructions were rewritten into equations. This operates neither with the concept of neurons receiving and sending signals, neither interconnect layers generating representations.

All of these innovations came from grasping the models in a conceptually new way and contain a little re-coding inside. However, none of the these recodings exhibits some predictive power in terms of experiments results, neither does provide conceptual framework that would cover all previous innovations.

In that way, deep learning is understood as a mathematical artifact, driven by the internal logic of mathematics (which is according to Kvasz of linguistic nature) and the social factors—from within the research community and the technological world—become secondary because the language of mathematics will find its way sooner or later. The technological spectacle about it has only little to do with it. As many times in the history (as was shown here on the example of trigonometric functions and many other examples by Kvasz (2008)), the current language of mathematics created objects (deep learning models) which cannot fully accommodate and the only way out is to come up with a higher level language that will be able to explain them fully.

Chapter 4

Methodology of the Empirical Study

This chapter introduces methodology, I later apply in chapters 5 and 6. In the first section of this chapter, I identify the relevant social groups and material I will work with further. The second section introduces the computational methodology of preprocessing of the data and getting a basic quantitative insight. The last section of this chapter summarizes the qualitative methodology I use to analyze selected texts.

4.1 Identifying Social Groups and Their Media

The term of relevant social group is the crucial one for the SCOT analysis. The definition says these are all social groups whose attitudes and opinions affect the emergence of technologies, however does not specify the criteria delimiting such groups in the society.

The presented classification of the groups and the list of online media targeted at them is based on a preliminary research which I would call hyperlink-snowball. [1] I started with well-knows web pages (Guardian, TechCrunch, Wired) and followed the links they referenced as source of the technology news. Then I searched the relevant discussion fora (StackOverlow, Reddit) for links to blogs. Skimming through the web pages helped me both to set

[1]This a reference to a name of a method of qualitative research in social sciences where researchers reach other test persons by receiving reference from the previous ones until they get references to the persons they already talked with. (Goodman, 1961)

the definition of the relevant social groups and find material for the further research.

I introduce four social groups based on the business and social role of their members. The relevant social groups are (ordered by the strength of technical insight to the technology):

- *members of the public*—people that do not have any special relationship to the technology, but are inevitably users of technology and their lives are affected by technology they use that is about to get available;
- *technology fans*—members of the public who are interested in technology development, follow recent trends and try to be progressive in use of technology;
- *IT specialists*—professionals whose main responsibilities are developing new products or technologies using programming languages and existing theoretical concepts, IT management or administration;
- *AI researchers*—professionals who invent, develop and empirically verify theoretical concepts and techniques in the field of AI.

The reason for choosing this classification is that the social and business roles defining the groups form a prototypical reader of the online media I work with. Membership in the groups is of course non-exclusive. Members of the last groups are likely to be technological fans which are all also members of the public. AI researchers often have experience from working as IT specialists.

In the following paragraphs, I will list the media I work with and briefly introduce each of them.

Public

Technology news for people with no special interest in technology can be found on common news servers. Those people of course may not read these news stories (and most their information about technology can come form their better-informed friends), nevertheless the articles themselves need to be written in such a way that is suitable for these readers. I have chosen the electronic versions of the well-established English-language newspapers.

- *The Guardian* (`https://www.theguardian.com/uk/technology`), and
- *The New York Times* (`https://www.nytimes.com/section/technology`).

Technology development, especially in case of emerging technologies is not usually a subject of political discussions and neither is subject of ideologically

	Twitter followers	Facebook likes	Feedly subscribers
TechCrunch	9.7M	2.7M	909k
Wired.com	9.7M	2.6M	734k
Gizmodo	2.6M	1.5M	701k
Ars Technica	1.2M	382k	365k
Mashable	682k	773k	10k
TechRepublic.com	201k	906k	—
Geek.com	93k	649k	27k

Table 4.1: The most popular technology news servers on Twitter, Facebook and Feedly as of September 2017.

motivated commentaries. Therefore, there is no reason to expect that political or demographic orientation of the newspaper can influence the way they inform about the technology. The audience of The Guardian and New York Times is considered demanding and well-educated, I can expect technical accuracy that will make searching for relevant articles easier.

Articles from the newspapers are included in the qualitative analysis only because their content is protected from massive downloading and thus cannot be used for the large-scale automatic analysis.

Technology Fans

By 'technology fans', I mean people whose interest in technology lead to active search for news about recent technology development. Although, many of those people are probably IT experts or have thorough education in other technical fields, the servers does not make any requirements for the reader in this sense.

The servers I have selected are the most popular technology servers in the world in terms of followers at Twitter which is the most important social network for news services. In addition, I take in account number Facebook likes and number of subscribers on Feedly[2]. Feedly is a web and mobile application that regularly informs its users about new articles on webs the users decide to follow. Particular statistics are tabulated in Table 4.1. Moreover, similarly to the previous media category, I do not expect the values attributed to the technology to differ much among media targeted to this group.

I decided to work with the following well-known major technology servers:

[2] `https://www.feedly.com`

- *TechCrunch* (`https://techcrunch.com`) — TechCrunch is an American online publisher of technology industry news founded in 2005. It publishes opinions on new products, thorough analysis of emerging trends in tech. It claims it has 6.5 million monthly active users worldwide, out of which 86 % allegedly declare technology as a personal fashion;
- *Wired.com* (`http://www.wired.com`) — Wired.com is an online version of American and British technological magazine. The website also hosts technology blogs on topics in transportation, security, business, new products, video games, the programming, cameras, culture, etc. It claims to have 2.5 million monthly active users;
- *Ars Technica* (`http://arstechnica.com/`) — Ars Technica is a website covering news and opinions in technology, science, politics, and society, created in 1998 by a group American computer scientists. Since 2008 it is owned by Wired. It has a separate British and American version having in total 680 thousand monthly active users.

The third most popular technology news server *Gizmodo* is actively preventing machine-processing of its content, there was not included in the study.

IT Specialists

IT specialists which are not experts in AI are potential direct users of the AI research. This is the reason why AI might appear on software developers' blogs among other topics which are of imminent relevance for their everyday jobs.

IT specialists and especially software developers tend to share various practical information on their blogs—best practices both in terms of work organization and the actual software development. They share experience with particular programming methodologies, new software libraries and development tools. Most of the blogs is targeted almost exclusively to relative closed communities of users of particular programming languages, software libraries and software development tools. For an outsider in most of the developer communities, it is difficult to estimate which blogs are the most relevant ones. Except for few well-known institutional blog, I rely mostly on the number of subscribes on Feedly (see Table 4.2).

The servers and blogs I decided to work with are the following:

	Feedly subscribers	Twitter followers
Coding Horror	124k	223k
Hacker News	113k	410k
Joel on Software	70k	151k
Scott Hanselman's Blog	66k	195k
The Daily WTF	62k	1.2k
Martin Fowler's Blog	57k	225k
Google Developers Blog	52k	—
High Scalability	47k	0.6k
The GitHub Blog	30k	—
David Walsh's Blog	29k	54k
good coders code, great reuse	11k	128k
OdeToCode	10k	39k
Virtuous Code	4k	18k

Table 4.2: Popularity of blogs targeted on IT professionals on Feedly and the number of Twitter followers in case there is Twitter account that can be directly associated with the author of the blog or the institution as of September 2017.

- *Coding Horror* (`https://blog.codinghorror.com/`) — a programming blog by Jeff Aftwood, a globally influential programmer, co-founder of StackOverflow, a community help server for developers with more than 10 million active users;
- *Hacker News* (`http://thehackernews.com/`) — server focusing mostly on system security targeted at IT specialist working as system administrators;for
- *Joel on Software* (`http://joelonsoftware.com/`) — blog about software development by Joel Spolsky, another co-founder of StackOverflow and an influential developer;
- *Scott Hanselman's Blog* (`https://www.hanselman.com/blog`) — personal blog of Scott Hanselman, a senior software engineer and manager at Microsoft; his blog posts focus on programming and so called smart devices;
- *The Daily WTF* (`https://thedailywtf.com/`) — it is a humorous blog calling itself a "software engineering disaster blog" running since 2004 which regularly publish the worse anti-patterns observed by hundreds of contributors in their software engineering practice;
- *Martin Fowler's Blog* — a personal blog of Martin Fowler, a British software developer, author and international public speaker on software development;

- *Google Developers Blog* (`https://developers.googleblog.com/`) — official Google's blog, contains post from Google employees mostly about Google technologies available for the developer community;
- *High Scalability* (`http://highscalability.com/`) — personal blog of Todd Hoff a former software engineer in major american technology companies (including Yahoo and IBM), now working as a free-lance consultant specializing on web services designed to work under extreme load; this is also the topic of most of the blog posts;
- *The GitHub Blog* (`https://github.com/blog`) — institutional blog of GitHub, web-based Git version control repository hosting service with over 60 million open-source projects with 20 million registered users worldwide; blog presents innovations on GitHub and best practices for open-source software development;
- *David Walsh's Blog* (`https://davidwalsh.name/`) — a personal blog of David Walsh, a web developer and software engineer working at Mozilla Foundation; his blog posts are focused mostly on web applications development;
- *good coders code, great reuse* (`http://www.catonmat.net`) — a personal blog of Peter Krumins posting about hacking, computer security and software development in general;
- *OdeToCode* (`http://odetocode.com`) — an influential blog about web programming written by K. Scott Allen, a developer and author of programming textbooks;
- *Virtuous Code* (`http://www.virtuouscode.com/`) — a blog by Avdi Grimm mostly about web programming.

AI researchers

For AI researchers, the main (and the most formal) means of communications are research papers, most of them published online as so-called *pre-prints* on arXiv (Steele, 2012). These are often contain small empirical results which would be unsuitable for a "proper publication", which are then interpreted by the community quickly. The recent results are then discussed on Twitter or the researchers' blogs. Unlike the actual research papers, which contain a lot of formal mathematical descriptions and formulas, the informal nature of the blogs, makes them easily computationally analyzed among the other texts.

Deep learning is more empirical than theoretical filed of study. The first-hand experience of deep learning practitioners is therefore more valuable here and this may be why they tend to share this experience with others on their personal or community blogs.

Unlike the previous categories which include also professional media who publish articles on regular basis, here I use exclusively personal blogs where the authors post only once in few weeks. Therefore, the number of sources is much higher than in the previous groups, however, the volume of text that can be collected is lower.

Whereas in case of the other groups, my goal was to choose a representative sample of the servers or blogs, in this category, I present what I believe is an exhaustive list of research blogs focusing on deep learning. On top of these, there are also few blogs with only one or two post where the authors stooped to contribute very early. I include only those who blog regularly for at least only year of the studied period. The popularity of the blogs on Feedly and their authors on Twitter is tabulated in Table 4.3.

The blogs I have chosen to work with are:

- *Google Research Blog* (`https://research.googleblog.com`) — a blog of Google Research presenting successes the company's research, most of them them are about machine learning;
- *No Free Hunch* (`http://blog.kaggle.com/`) — institutional Blog of Kaggle, Google-owned company organizing both commercial and educational contests in machine learning; blog announces news competitions and comments on results and method used by the participants (its name is an allusion to a famous mathematical theorem regarding machine learning which is colloquially called 'No free lunch theorem');
- *Machined Learnings* (`http://www.machinedlearnings.com/`) — a personal blog of Paul Mineiro, a researcher working at Microsoft Research; contains mostly comments on recently published papers;
- *NLPer's Blog* (`http://nlpers.blogspot.com`) — blog of Hal Daumé III., a prominent researcher in natural language processing at University of Maryland discussing scientific topics and topics concerning academic life;
- *Daniel Lemire's Blog* (`https://lemire.me`) — Daniel Lemire is a computer science professor at the University of Quebec; the topics oscillate between

technology oriented comments on deep learning applications and comments on purely scientific topics;

- *Hacker's Guide to Neural Networks* (http://karpathy.github.io/) — a blog written Andrej Karpathy, world's leading researcher in AI where he explains some of his papers and shares his opinions on AI development; started in 2011;
- *Colah's blog* (http://colah.github.io/) — blog with intuitive and better understandable explanation of recent DL innovations;
- *Cortana Intelligence Blog* (https://blogs.technet.microsoft.com/machinelearning/) — a blog run by Microsoft Research, presents both newest academic achievements of Microsoft Research and launches of new Microsoft cloud services using deep learning;
- *Calculated Content* (https://charlesmartin14.wordpress.com/) — a blog by Charles Martin, free-lance machine learning consultant, started in 2012, math-oriented;
- *Tim Dettmer's Blog* (https://timdettmers.com) — blog a of PhD student from University of Lugano; the declared goal of the blog is to make deep learning accessible to everyone; the blog is also concerned with organization of the scientific life;
- *Foldl* (https://fold.me) — blog of Jon Gauthier, a PhD student at MIT; in his blog posts, he focuses mainly on using deep learning for natural language processing.

4.2 Filtering and Qualitative Analysis of Relevant Media

This section describes the techniques that were used to obtain the data and how these were further processed and analyzed. Source code of all the processing steps is publicly available on my GitHub profile (TBD).

Because the Internet is an ever-changing environment, instead of crawling the pages directly from their URLs, I use their archived versions as captured by *Wayback Machine* project of the Internet Archive Foundation (https://web.archive.org). The goal of the project is to archive the content of the web for further reference and research.

	Feedly subscribers	Twitter followers
Google Research Blog	39k	—
No Fee Hunch	18k	90k
Machined Learnings	8k	0.7k
NLPer's Blog	6k	11k
Daniel Lemire's Blog	6k	6.7k
Hacker's Guide to Neural Networks	4k	93k
Colah's blog	3k	14k
Cortana Intelligence Blog	3k	—
Calculated Content	1k	—
Tim Dettmer's Blog	392	1.7k
Foldl	146	1.9k

Table 4.3: Popularity of blogs on AI topics on Feedly and the number of Twitter followers in case there is Twitter account that can be directly associated with the author of the blog or the institution as of September 2017.

From the content of the downloaded websites, I extract plain text which later undergoes computational linguistic analysis. In particular, this is automatic morphological and syntactic analysis, which is used to filter out word with grammatical meaning only, and entity recognition, which prevents excluding multi-word terms.

In the next step of the automatic analysis, I extract keywords using the *tf-idf* algorithm. This algorithm assigns every term in a document a score capturing how descriptive the term is for the document. The score is computed as a product of the term frequency in the document (the *tf* part) and a logarithm of the inverse document frequency (the *idf* part). The fewer documents contain a specific term, the higher *idf* the term gets. For uncommon terms, it suffices to appear few times in a document to get a high score. On the other hand, common terms need to appear frequently to get a high score.

The extracted keywords are used to select relevant documents which are analyzed further using *Latent Dirichlet Allocation* (LDA) (Blei, Ng, & Jordan, 2003).[3] Currently, this is probably the most commonly used method for automatic topic analysis.

[3]This method therefore falls into the previous paradigm in machine learning. It would be nice to analyze deep learning using deep learning, however topic modeling remains on of the areas of natural language processing that is still done better using Bayesian statistics, nevertheless some attempts to introduce deep learning into topic modeling already exist (Moody, 2016).

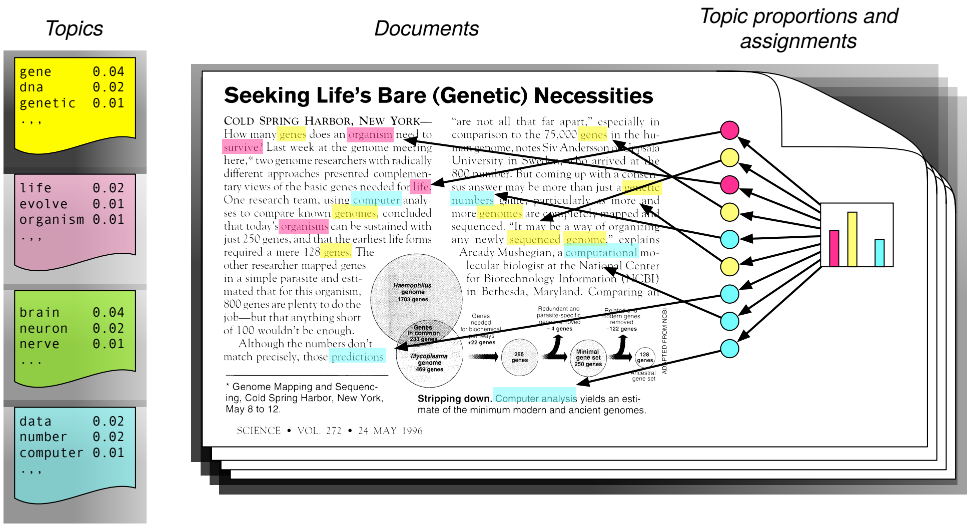


Figure 4.1: Illustration of the probabilistic mechanism used by the LDA algorithm. The example is taken from ICML 2012 Tutorial slides (Blei, 2012).

Both keyword extraction and LDA view the texts as a set of terms disregarding their order. These terms are usually content words from the texts (words having only grammatical meaning as pronouns or prepositions are removed), often the words are also lemmatized (transformed to a basic form). Proper names and frequently used multi-word expressions (in linguistic literature called *collocations*) are treated as if they were single terms.

LDA assumes that each document (i.e., a set of words) is associated with a probability distribution of topics. Each topic generates different terms with different probabilities. This simplification further assumes the document is created using a stochastic process which works in the following way: every time we want to add a term into the document, we randomly select one of its topic (proportionally to their probability). From this topic, we randomly select a term (again proportionally to the term probabilities within the selected topic).

If we already know the topics, we can assign any document with a probability of the topics that were likely to generate the document. The topics (as distribution of words) can be estimated from big corpus of training text. The method works in an unsupervised fashion, i.e., there is no need to decide about the topics beforehand, the only thing that needs to be decided in advance is the number of topics. A semantic label for the set must be however assigned manually. The method is illustrated in Figure 4.1.

I used *html2text*[4] to extract plain text from downloaded websites. The text were then processed using natural language processing library *spaCy*[5]. The LDA is performed using the Gensim library[6] (Řehůřek & Sojka, 2010).

Results of these methods and their interpretation and analysis is presented in Chapter 5.

4.3 Discourse Analysis

Because all the studied articles are targeted to different audiences—social groups relevant for the technology—I can also assume that there is a separate instance of discourse for each of the categories. This observation encourages me to critically study the language of the articles with an approach that could be assigned a label of *Critical Discourse Analysis* (Fairclough, 1995). The adjective *critical* refers to the attempt to discover ideologies and power relations involved in discourse. Discovering the value system and the underlying ideology which motivates behavior of relevant social groups towards the deep learning technology is after all the main goal of this thesis.

Discourse analysis often works with language in a big detail, analyzing morphology, syntactic means the authors use or lexical choice (Given, 2008, p. 146). This detail-oriented level analysis would be inappropriate here. Often, the authors of the articles and blog posts are not native speakers of English (and neither am I), so there is a big chance of misinterpretation.

What, however, could be easily analyzed is the role that the actors (companies, researchers, users, etc.) play in the text: who is in an active and who in a passive position. Interesting insight can be obtained through a more general semantic properties of lexical means used in the articles.

For each of the categories I carefully analyzed tens of articles from which I have selected around 10 articles which represent the content of the articles the best. In case of the articles that have been automatically downloaded, the selection was based on 50 randomly selected articles in order to preserve the distribution of covered topics and also distribution in time.

Results and analysis of this method is presented in Chapter 6.

[4]`http://www.aaronsw.com/2002/html2text/`
[5]`https://spacy.io`
[6]`https://radimrehurek.com/gensim/index.html`

Chapter 5

Findings of the Quantitative Analysis

This chapter presents results of the computational analysis performed on the articles from the selected online media. In the first section, I describe the acquired data and the process of their cleaning and pre-processing. The following two sections present results of keyword and topic analysis and comment on them. The chapter ends with preliminary conclusions that can be drawn from the analysis results.

5.1 Acquired Data

In total, I downloaded 33GiB of websites (only HTML data without images, videos, PDF files, etc.) from which I was able to extract 1.4GiB of plain text. The data contains almost 84 million words split in ratio 180:4:1 into websites targeted to technology fans, IT specialists and researchers. A detailed overview for the individual websites can be found in Table 5.1. The news servers for broad public are not included in the computational analysis because their content cannot be easily automatically downloaded and it is explicitly prohibited by the server operators to do so.

In case of developer and researchers blogs, a relatively big proportion of articles had to be discarded because it was impossible to extract coherent text from them. It was usually caused by the fact that many of the pages contains mostly some source code and the accompanying text is in a form of short commentaries in the code. In case of the servers for technology fans, the majority of discarded pages were video post withh only a short textual commentary.

website	documents	words	keep ratio
Ars Technica	10,411	8,076,158	51%
TechCruch	60,456	34,002,742	95%
Wired.com	44,375	39,671,054	38%
Technology fans	111,242	81,749,954	93%
Coding Horror	83	107,884	96%
Google Developers Blog	272	128,470	97%
OdeToCode	312	150,633	87%
The GitHub Blog	592	227,438	94%
David Walsh's Blog	460	337,702	86%
High Scalability	248	519,215	42%
Martin Fowler's Blog	50	222,919	100%
The Daily WTF	920	628,847	78%
The Hacker News	3,551	1,621,101	91%
Joel on Software	13	14,977	81%
good coders code, great reuse	113	62,646	81%
David Walsh's Blog	650	568,192	99%
Virtuous Code	37	35,362	57%
Developers	7,091	4,625,386	83%
Google Research Blog	190	127,718	97%
No Free Hunch	300	362,160	92%
NLPer's Blog	45	47,588	94%
Daniel Lemire's Blog	290	207,348	95%
Hacker's Guide to Neural Network	14	48,455	67%
Colah's Blog	13	38,603	100%
Cortana Intelligence Blog	124	115,667	90%
Calculated Content	34	47,896	18%
Tim Dettmer's Blog	7	32,259	44%
Foldl	28	28,486	78%
Researchers	1,519	897,597	82%
Total	119,852	86,912,947	92%

Table 5.1: Overview of the amount of the data extracted from the web pages.

5.2 Keyword Analysis

For the keyword extraction, I used the *tf-idf* algorithm that assigns a score to every term in an article. The absolute values of the scores are not interpretable outside of the documens scope. It is the rank of the terms which should be used to decide whether a term should considered a keyword in the articles. For further processing, I have selected documents when one of the following key phrases: *deep learning*, *artificial intelligence*, *machine learning*, appeared the among the first 100 best-scoring terms. The presence of a keyword among the best-scoring ones can be also used to analyze how frequently relevant articles appeared during the studied period.

Figure 5.1 shows the time lines of the selected key phrases occurrence in the 100 best-scoring keywords during the observed period. Even from this basic analysis we can observe several interesting facts.

In the articles targeted on the technology fans, the term is the most frequent one *artificial intelligence* over the followed terms. It was of little use before 2013, and its popularity starts to grow together with *deep learning* in a similar pace as *machine learning*. Proportion of the articles on AI on server for technology fans increased from almost none in 2012 to 4% in 2016.

The proportion of articles and blog posts about AI and deep learning targeted on IT specialist is surprisingly low, almost 2 times smaller than in case of articles targeted on technology fans. From this I hypothesize that software developers do not consider the research in AI to be their field at all. Even in case when the technology that they uses machine learning in the backend, it is trifling to them.

The blog post targeted on researchers tend not to call the *deep learning* models as AI. The peak of the usage frequency of *machine learning* is probably connected with the previous generation of machine learning methods. While taking into consideration that *machine learning* and *deep learning* are used merely as synonyms, we can see that more than a half of research blogs in AI-related research is about deep learning. The proportion of the articles containing selected keyword grows to almost 50% during the time.

For the further processing, I selected only the articles where one of the selected keywords scored among the top 100. The proportion of relevant articles on servers for technology fans and IT specialists is shown in Table 5.2. I use all articles from researchers' blogs because all of them are relevant to the

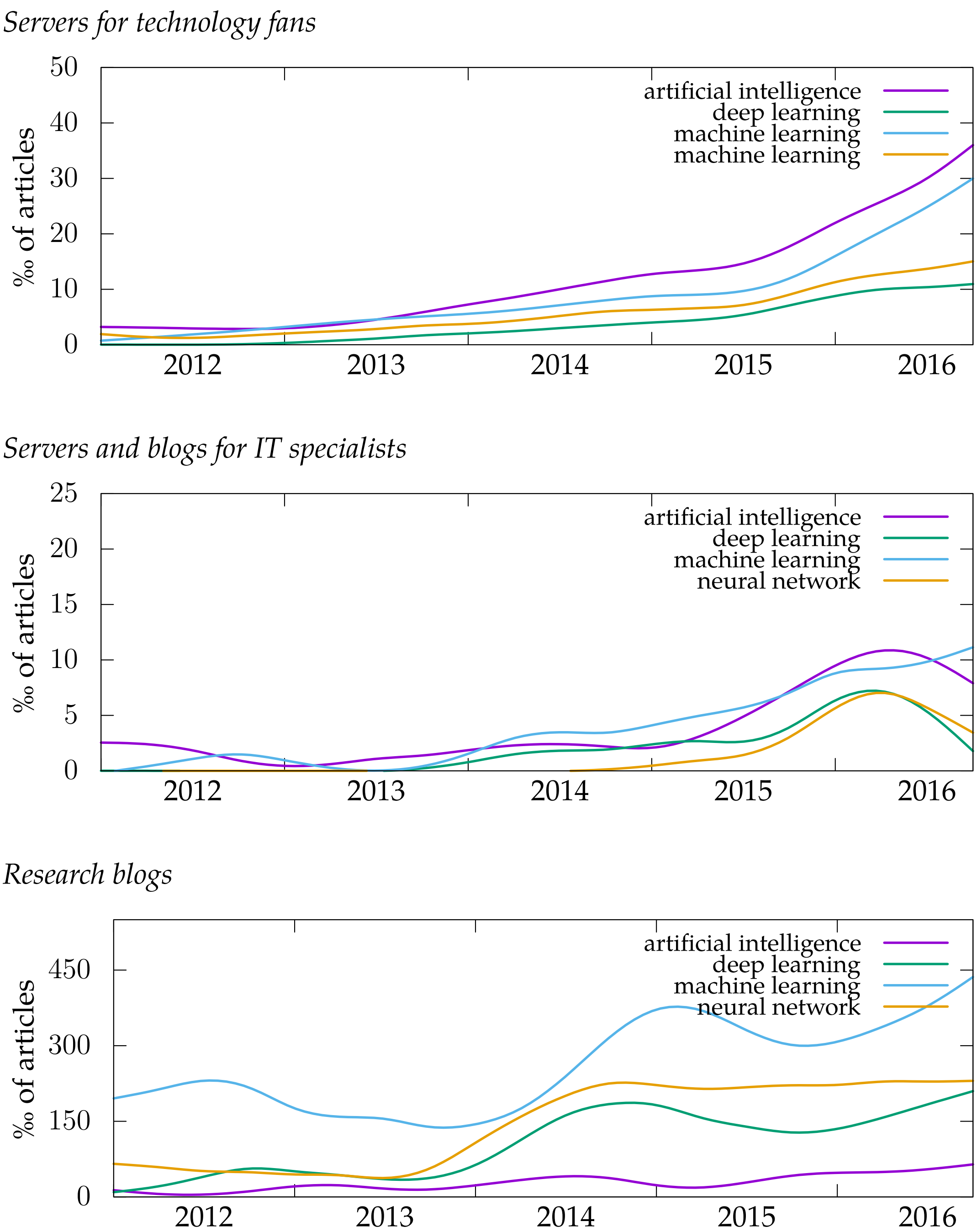


Figure 5.1: Relative frequency of articles (in per mille) where the keywords of interest appear among the 100 best-scoring keywords. Values are computed per quarter, curves are smoothed. Note the y axes are in different scales.

topic, even though they may not contain the mentioned keywords. They often work with more technical terms (like *representation learning*, *convolutional network*, *sequence-to-sequence learning*, etc.). This may raise concerns whether this is not also the case of the articles for IT specialists. Nonetheless, these terms are too specialized for someone who is supposed to be only a user of the technology. This also confirmed while reading through the articles.

website	documents	proportion of data
Ars Technica	45	0.4%
TechCruch	1,413	2.3%
Wired.com	655	1.5%
Tehcnology fans	2,113	1.9%
Codding Horror	2	2.4%
The Hacker News	20	0.5%
Scott Hanselman's Blog	2	0.3%
Google Developers Blog	7	2.5%
High Scalability	12	4.8%
The GitHub Blog	2	0.3%
David Walsh's Blog	1	0.2%
good coders code, great reuse	5	4.4%
Developers	78	1.1%

Table 5.2: Number of articles from various sources selected for further analysis (one of the tracked keywords appeared among the 100 highest scoring ones).

5.3 Topic Analysis

Unlike the isolated keyword analysis presented in the previous section, which works with the terms in isolation, the topic analysis discovers clusters of related terms and their possible dependencies. Both 2,113 relevant articles and all 111,242 targeted to the technology fans are analyzed to make a comparison in the topic structure. Servers targeted on the developers are omitted due to the lack of relevant articles. On top of 1,519 articles from research blogs, I also performed the topic analysis on abstracts of 19,100 research papers from the AI category published on *arXiv.org* during the same period of time. Note that papers on *arXiv.org* are pre-prints that did not necessarily undertake a peer review process. This means that the results are getting published quickly—the topics are up to date with the current technological trends—often counterbalanced by the lower quality of the papers.

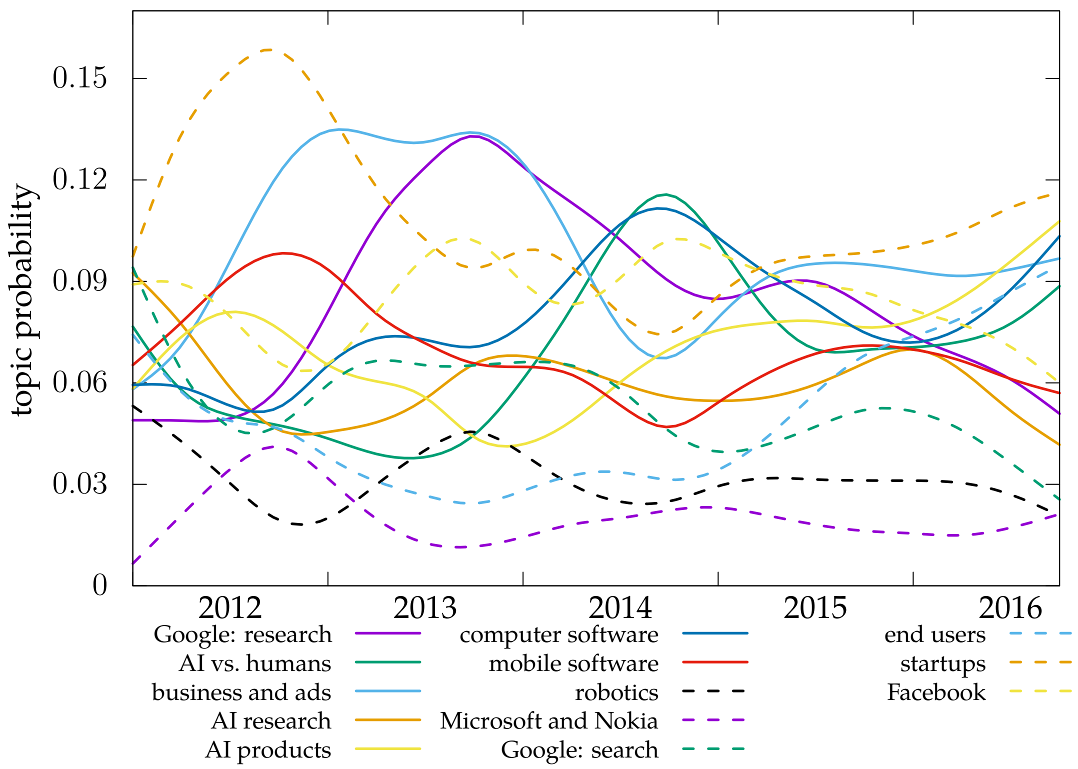
The development of the topics on servers for technology fans in time is shown in Figure 5.2. The most prominent topics are mobile applications and social networks. The other topic which stands out is science news. Despite this relatively high coverage of scientific topics, articles about deep learning and AI do not stress the research perspective on the development. The most important topics are start-up companies and related business topics. The increasing importance of articles about end-user experience is probably due to high proportion of software related topics in general and increasing role of deep-learning based technologies in the applications.

In 2013, research done by Google was among the highest scoring topic. The later drop of importance of this topic can be explained both by considering deep learning to be more a technology than a research topic. It also goes together with decrease of interest in Google that can be observed in all articles.

In the last year, there was also an increased number of articles comparing the deep-learning-based technologies and human abilities. This might be interpreted as a growing interest in the abilities and potential impact of the technology. Interestingly, this topic appeared only after the applications started to be deployed and were not discussed much before.

A totally different topic structure appears on research blogs as can be seen in the most bottom plot of Figure 5.3. Whereas in 2012, the most important topic was the scientifically most interesting topic of representation learning (the ability of neural networks to learn a vector representation of its inputs which has many interesting properties), later more technical topics started to dominate. In particular, these were the models architectures and technical details of the model training. In 2016, the importance for machine vision, natural language processing and application in general increased. This probably reflects the increasing number of practical applications of deep learning. In general, research blogs tend to shift from scientific topics to more technological topics. Any non-technical aspects seem to be neglected entirely.

The topics found in the *arXiv.org* papers show only a slight increase in *natural language processing* and *machine vision*. Interestingly, deployment of application did not influence the popularity of topics much. I would expect much bigger increase in the popularity, so that the researchers could show off their participation on the most recently deployed technologies. On the other hand, *Relevant articles from servers for technology fans*



All articles from servers for technology fans

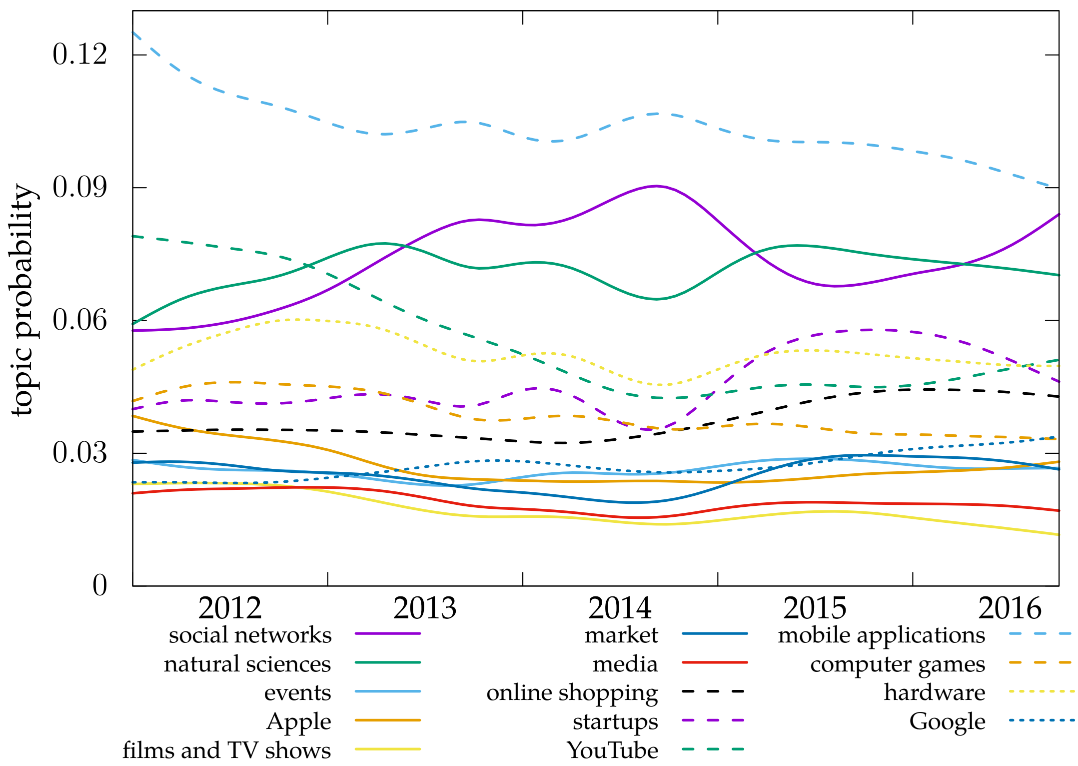


Figure 5.2: Development of LDA topic scores in time for articles for technology fans.

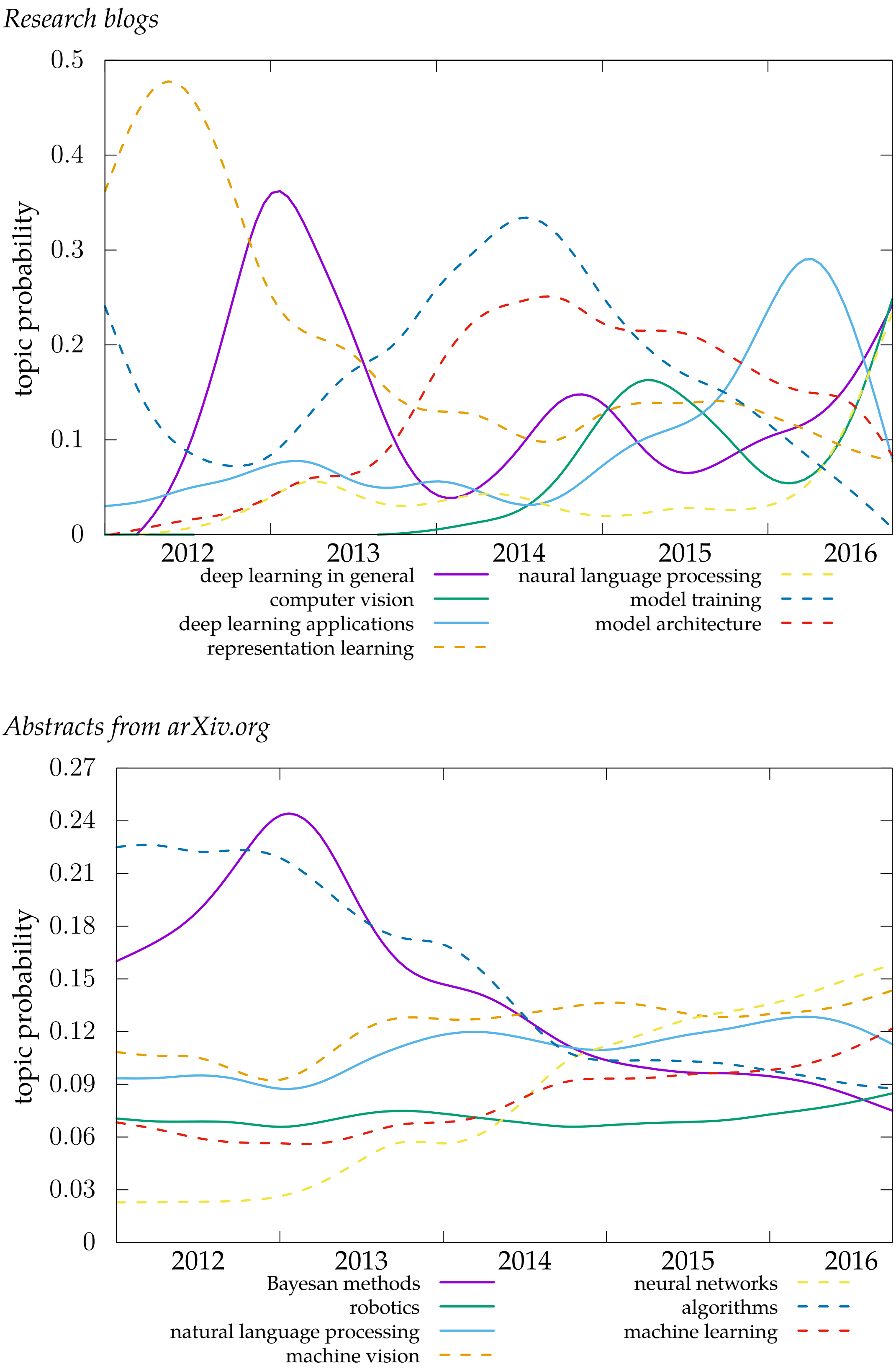


Figure 5.3: Development of LDA topic scores in time for research blogs compared to topic in abstract of papers from arXiv.org.

popularity of *neural networks* increased almost 5 times during the monitored period. An interesting moment is a peak of paper on Bayesian methods appears in the same time when researchers started to write posts about deep learning, whereas the blog posts do not address this topic at all. I hypothesize this is because the researchers wanted to publish their results achieved within the previous paradigm at least as pre-prints before they become outdated. This shows that the researchers blogs in some sense sets or at lest predicts trends that show in the research papers later on.

5.4 Preliminary Conclusions

The computational analysis showed many interesting patterns in the downloaded articles and blog post. From the presented analysis of the results, I can draw the following preliminary conclusions that are explored in more details in the qualitative analysis in the next chapter:

- The interest of public grows with number of technological applications of deep learning and tend call the solution AI.
- Servers targeted on technology fans connect AI and deep learning more with products that enter the market (their readers are likely to be early adopters) than with scientific research.
- IT specialists seem to be agnostic to research in AI, even though some of the technology originates in AI research.
- The research blogs shift from research topics to technological topics.
- AI researchers tend to avoid the term *artificial intelligence* even in their informal blog post.
- Research blogs anticipate topics that later appear in research papers.

Chapter 6

Findings of the Discourse Analysis

In this chapter, I present results of the qualitative discourse analysis done on selected articles on deep learning targeted on different audiences. The sections of this chapter cover news servers and blogs for audiences described in Chapter 4: general public, technology fans, IT specialists and AI researchers. At the beginning of each section, I list around ten articles that illustrate the best the observations that I am discussing underneath. The selected articles are always listed at the beginning of the section and are referenced with their alphanumeric identifiers thorough the rest of the sections.

6.1 Public

Selected articles

P1 Guardian, 3.10. 2012: Philosophy will be the key that unlocks artificial intelligence.

P2 NY Times, 23.11. 2012: Scientists see promise in deep-learning programs.

P3 Guardian, 21.3. 2014: Zuckerberg and Musk back software startup that mimics human learning.

P4 NY Times, 15.2. 2016: Google's computer program beats Lee Se-dol in Go tournament.

P5 Guardian, 15.3. 2016: AlphaGo: Beating humans is one thing but to really succeed AI must work with them.

P6 NY Times, 28.6. 2016: The promise of artificial intelligence unfolds in small steps.

P7 Guardian, 30.8. 2016: Google DeepMind and UCLH collaborate on AI-based radiotherapy treatment.

P8 Guardian, 13.10. 2016: Why workers needn't fear the new robot age.

P9 NY Times, 14.12. 2016: The Great A.I. Awakening—How Google used artificial intelligence to transform Google Translate, one of its more popular services — and how machine learning is poised to reinvent computing itself.

Observations

The articles about deep learning targeted on public have usually a similar structure. They begin with claims that dreams of artificial intelligence are already coming true, mention the (outdated and from today's perspective irrelevant) biological motivation. This makes an impression that the presented technology is something that goes down to the roots of human intelligence (to be accurate they mix up what Searl calls strong and weak AI, discussed in Section 1.1). This is often followed by comparisons of number of neurons in human and animal brains and in a deep neural network giving an unspoken hint that something as outrageous as a brain simulation might be going on. This is then followed by presentation of fascinating results that the technology achieved (flawless speech recognition, human-level machine translations, super-human performance in the game of Go). The articles look more similar to those reporting on scientific news. However, they are always a mix of scientific endeavor and enthusiasm on technology entering the market.

In many of the articles, there is also a lonely genius stereotype (P2, P6, P9) who in this case is Geoffrey Hinton (who certainly is a genius and his work deserves respect and admiration without any doubts). The stereotyped story can be summed up as follows: having been despised by the community for long years, he insisted on his vision and at the end it showed up he was a genius.

This myth tells the readers that deep learning is a big invention comparable to those made by other lonely geniuses from the history like Copernicus or Albert Einstein who must have experienced the same disdain before they finally got recognition. The myth also contains reassurance that there is nothing to worry about. Hinton is not doctor Frankenstein because the lonely genius myth got completed—Hinton got the recognition he deserves. Frankenstein did not get any recognition, his story is the exactly opposite one: he started as an esteemed scientists who eventually ended up being despised. Acknowledging the myth here matches deep learning and artificial intelligence with the story of never-ending technological progress. This reassurance is probably there to eliminate the disturbing feeling that was provoked by the unjustified implicit claims about brain simulations.

An article from 2012 (P1) comments the state of the art in AI research at that time and does mention deep learning at all. It references science fiction and most importantly, does connect AI with any existing or expected technological products, AI is a distant future in the article.

An article from similar time (P2) announces first successes of deep learning, mostly in academic competitions. Many positive emotional adjectives are used to describe the results (stunning, impressive, breathtaking). AI research is here no longer connected with science fiction but with a fuzzy prospect of new products.

In 2014 (P3), the news informed about investments into AI startups made by big technological companies during that period. The outlook of new products is still fuzzy, however emotional evaluation of deep learning success is no longer there. Possibly, because it already became a business, instead of a toy of masters students. The enthusiastic rhetoric is on the other hand present in articles commenting the success of AlphaGO, a program beating the world's best player in a game of Go (P4, P5).

In the last months of 2016, deep-learning-based technology received relatively big attention (P2, P7, P8, P9). None of the articles forgets to mention the biological inspiration of neural networks that still gives the technology a sci-fi impression. Deep learning is described as something that delivers excellent results. The products (probably due to the marketing effort of the technological companies) are always such that no one can doubt their benefits for the user or whole society (health care, better traffic organization, making life easier). Controversial topics like use of web search logs to improve advertisement targeting are silently omitted.

An exception to the product-oriented articles is a so called long read from NY Times called 'The Great AI Awakening' (P9) which tries to narrate a story of a fascinating scientific and technological adventure that eventually led to launch of deep-learning models in Google Translate. The article uses all elements that an adventure story should have: lonely genius hired by Google inspired others to an endeavor whose story is remarkably similar to documentaries on space missions.

In all the articles, users of the technology are attributed a passive role—their lives are promised to improve via a technology that comes from the labs of the world's best universities and technology companies and will be delivered to their smart phones and computers without any effort from their side. They should be excited about it as if opting out was out of consideration.

6.2 Technology Fans

Selected articles

- F1 TechCrunch, 23.10. 2013: Yahoo acquires startup LookFlow to work on Flickr and 'deep learning'.
- F2 Wired, 9.12. 2013: Facebook taps 'deep learning' giant for new AI lab.
- F3 TechCrunch: 11.7. 2014: Salesforce buys big data startup RelateIQ for up to $390M.
- F4 TechCrunch, 18.1. 2015: Autonomous cars are closer than you think.
- F5 Wired, 9.4. 2015: Toyota finally gets serious about self-driving cars.
- F6 TechCrunch, 18.1. 2016: Learn deeply, but baby, don't fear the Skynet.
- F7 Wired, 22.3. 2016: Google photos now builds perfect vacation albums on its own.
- F8 TechCrunch, 15.3. 2016: Google AI beats Go world champion again to complete historic 4–1 series victory.
- F9 Wired, 19.5. 2016: What the AI behind AlphaGO can teach us about being human.
- F10 Wired, 28.10. 2016: How AI is shaking up the chip market.
- F11 Wired, 9.11. 2016: Trump's win isn't the death of data—it was flawed all along.

Observations

Unlike the articles for public which do not forget a brain-like magic behind the technology, articles targeted on technology fans refer about deep learning as if it were a technology behind new products any other. The terms in which the technology is described sometimes seems to be more suitable for Wall Street Journal than for a technology server.

This is probably related to the fact that the technology fans are early adopters of the technologies and try to pose themselves as trend-setters for people which are not so interested in technology. People self-confidently moving through the technological world obviously cannot let themselves get astonished by artificial intelligence. They want to make an impression that technology is something under their control, something they have observed for a long time and cannot bring any surprise to them. This is what the technology servers offer them. The way of getting the control is fitting everything into well-known categories of technological market.

Another reason for that might be that marketing of technology companies tries to make an impression that computer science is something that happens on the market (no matter whether we talk about market when the products or services are paid by users or by advertisers). As potential customers and influencers of other customers, technology fans need to be constantly reassured that the cool things are exactly those they can buy or subscribe to. Most importantly, they need to show that there is no need to seek them within some hacker communities promoting open-source movement and pirate ideology.

In the first three years of the monitored period, the mentions about the products are as fuzzy as in case of articles for public (F1, F2, F3). However, deep learning gets never legitimized via its academic success, but always via its potential market value. Big companies are willing to invest a lot of money, so it must be important. This logic is even more evident in case of later articles that talk in more detail about particular products (F4, F5, F7, F10).

The presentation of deep learning is in contrast with how the servers report about science which receives a relatively high coverage as shown in the quantitative analysis (see Section 5.3). Unlike technology, science is something the technology fans can afford to get astonished about. The only exceptions are articles presenting surprisingly good results of AI—like the AlphaGo beating the world's best players (F8). However, the opinion article published few days afterwards (F9) already speculated about possible products and monetization of the technology.

Some articles (F11, F6, F7, F10) react to unspoken worries about privacy issues connected with processing of big data and potential market distortions due to availability of both the know-how and data only to some market players. These worries are never argued carefully, however they are always quickly disapproved as unjustified. The worries are explained as raised by outsiders, people who do not know much about technology. Their concerns are presented as understandable—after all there is a danger of misuse in every new technology. However, the insiders among which the audience of the servers is automatically counted, know much more than the concerned outsiders.

6.3 IT Specialists

Selected articles

I1 The Hacker News, 25.6. 2015. Facebook can recognize you even if you don't show your face.

I2 The Hacker News, 9.12. 2015. It works! Google's quantum computer is '100 million times faster' than a PC.

I3 Google Developers Blog, 26.1. 2016: Teach yourself deep learning with TensorFlow and Udacity.

I4 The Hacker News, 19.5. 2016: Hey Allo! Meet Google's AI-powered smart messaging app.

I5 Google Developers Blog, 15.12. 2016: Start with a line, let the planet complete the picture.

Observations

The automatic analysis of the texts targeted on developers and IT specialists already showed that when they act in a role of an IT specialist (most of them certainly belong also to the group of technology fans and in that role they probably do care), deep learning is out of their interest. This is the case even when the article introduces a service with deep learning models in the back-end.

The only exception is the Google Developers blog whose goal is to promote and introduce Google's technologies available for developers (I3, I5). It is obviously a part of Google's marketing strategy to let everybody know that the company mastered this advanced technology.

There were also articles in The Hacker News that mention deep learning (I1, I2, I4). However, these articles are more similar to articles from TechCrunch than to the rest of Hacker News focusing on technical details of IT administration.

The reason why there is a little coverage of deep learning in articles for IT specialists is probably that deep learning is not an essential part of software development in general. Moreover, mastering deep learning requires a big effort that not every IT specialist is willing or can afford to spend. Those who are, probably search for more specialized resources.

This can make a false impression that IT specialist are not a relevant social group for deep learning at all. This not true, their silence is significant. They are the primary users of the technology. Developers are the people that turn the technology into products. Their agnosticism to the underlying principles of the technologies and reluctance to attribute a special status to deep learning (and probably also other complicated theories that are behind many software libraries), necessarily leads to seamless encapsulation of the technology while integrating into products. This may also make it easier to leave out both the worries and fascination by the self-learning ability of the AI algorithms and present the products as any innovation as in the articles for technology fans.

6.4 Research Blogs

Selected articles

R1 Hal Daumé's blog, 15.9. 2012: Somehow I totally missed NIPS workshops!

R2 Google Research Blog, 12.2. 2013: Research projects on Google App Engine.

R3 Google Research Blog, 22.7. 2014: Academics and the little box challenge.

R4 Colah's Blog, 13.7. 2014: Understanding convolutions.

R5 Google Research Blog, 15.4. 2015: Google handwriting input in 82 languages on your Android mobile device.

R6 Colah's Blog, 3.9. 2015: Neural networks, types, and functional programming.

R7 Andrej Karpathy's Blog, 31.5. 2016: Deep reinforcement learning: Pong from pixels.

R8 Google Research Blog, 29.6. 2016: Wide & deep learning: better together with TensorFlow.

R9 Hal Daumé's blog, 24.6. 2016: Language bias and black sheep.

Observations

Establishing deep learning as a technology more than a research topic, is also reflected in the blog posts by AI researchers. Most of the blog posts (except for Google Research Blog) are technical and pay attention to particular technical or mathematical aspects of deep learning. Surprisingly, the private blogs do not tend to comment emotionally on achieved results, they are technical not only in the content, but also in the language (R4, R6, R7, R9).

The reason (probably sometimes unconscious) for writing a research blog is getting recognition in the scientific community. The blog posts often try to explain in a less formal way what seems to be unclear from formal research papers, even though the language of the blogs is not what would public perceive as informal. On the other hand, it still makes a big difference from the strict style of research papers which leaves only a little space for the researchers' motivations or for illustrating metaphors which are often present in the blog posts (R4, R6, R7).

The serious tone and technical style might be a symptom of trying to show control over the topic. Experiments with deep learning are sometimes empirical (unlike natural sciences, it is often hard to say what the outcomes will be) which is something that does not belong to scientific discourse. Therefore, it is marginalized not only in the papers but sometimes also in the blog posts. Most of the blog posts thus try to show some kind of control over the performed experiments and in this way advocate the authors' membership in the scientific community.

Conclusions

Deep learning is an emerging technology standing behind many recent innovations in the IT world. During the studied period (2012–2016), it bridged the gap between a purely research technology and production deployment, while being widely adopted by all big technological companies.

Whereas in 2012, NY Times mentioned deep learning as a promising research program, in 2016 it was presented as a mature technology that is empowering the best products of technological companies. This shift is apparent in articles targeted on all four groups I have identified at the beginning. Ordered by their "distance from the technology", these are: AI researchers, IT specialists, technology fans and public. There is a wide range of online media (both traditional and new media) targeted on these groups. I analyzed articles and posts from these media and tried to recognize the values the relevant groups associate with the technology.

During the studied period, the researchers shifted their attention to practical questions regarding deep learning from topics that are more interesting from the research perspective which they were the most interested at the beginning. As the technology gets more widely adopted, the researchers do not want to lose the status of its founding fathers and therefore need to at least partially accept the values of the technology users. The blogs are usually written to show and to confirm the status of the authors in the community and often to demonstrate control the researchers have over the field. A way to show it, is talking about the topics everybody knows are related even though they may have been introduced by marketing activities of technological companies.

IT specialists seem not to pay any special attention to the technology (at least in the role of IT specialists, in a role of technology fans they probably do). Their silence is however constituting for deep learning as a technology. In order to be used by software developers, it has to be a software library like any other. Everything that is too specific for the technology needs to be encapsulated and moved out of the reach of the developers. They silently push the deep learning practitioners to make deep learning a backend technology from the developers' view indistinguishable from the others. This also makes it easier to present the technology similarly to the technology fans.

The technology fans do not want to make an impression of being overwhelmed by innovations enabled by deep learning. As early adopters of technologies they also act as opinion makers and influencers to their social surrounding and cannot afford to lose their status of having control over the technology development by attributing any technology a special status. Any concerns about privacy, possible negative social impact or unexpected uses of technology are attributed to the technology outsiders, whereas the insiders know there is nothing worry about. Admitting such concerns would only mean admitting not having the situation firmly in their hands. This at least the image that servers targeted on technology fans try to make, probably to some extent influenced by the marketing of technological companies.

Surprisingly, it is the public who shares with the researchers the fascination by the fact machines are performing interesting things via self-thought cognitive simulations. However, this is message is delivered to the public in quite a malformed shape. The original goal of AI research—study cognition and intelligence via its computational simulation has reduced to an anecdote about number of neurons in human, mouse or monkey brains without any meaning other than providing an illusion of exclusivity. The technology is explained like there was never a different goal other than to empower such praiseworthy application like automatic X-ray analysis or machine translation. These are mentioned probably to prevent any worries about unintended consequence of the technology introduction.

The story of deep learning is often narrated using a myth of a lonely genius, who despite long-lasting disdain by the scientific community finally received recognition. This myth implicitly puts deep learning to the same position as other instances of this myths like stories of Copernicus or Albert Einstein. Deep learning is often presented with unspoken positivist assumptions, probably even stronger than the researchers put into it. Comparisons with such mystical object as human brain only stress the unspoken program of taming nature via technological means.

The findings summarized in the paragraphs above fit well into the Bijker's framework of SCOT. What was originally seen as a research topic by the research community and presented in this way in the media targeted on public, was overtaken by technological companies which opened the technology to other relevant social groups. The public turned from distant observers into users of the technology. Standards and belief within the IT specialists community shaped the technology in such a way that it can treated the same way any other software technology.

The findings could also well explained using the concept of 'participation' (Carpentier, 2012) and maintaining power over the technology. This principle can be used both for explaining the topic shift in the research blogs from scientific to technology topics, the reluctance of IT specialist to attribute a special status to deep learning, and providing the technology fans an impression of having control over new technologies.

As far as I know, this thesis is the first study that identified relevant social groups for deep learning and analyzed media targeted on these groups. Nevertheless there were two recently published studies dealing with how AI is presented in the media.

Findings of a study analyzing New York Times articles on AI between 1985 and 2015 (Fast & Horvitz, 2017) are similar to the qualitative analysis I have done on similar data (news servers targeted on public). The study emphasis positive and enthusiastic rhetoric of the articles and expectations connected with AI application in medicine and education which was observed between years 2009 and 2015.

Another study conducted on tweets (Manikonda & Kambhampati, 2017) divides the authors of the tweets into two groups: experts and non-experts. From the examples presented in the paper, I expect the non-expert tweeters to the group of technology fan and the expert tweeters to the group of AI researchers. The authors focus mainly on estimating the sentiment of the tweets. Their rather shallow conclusions are that tweets on AI are mostly optimistic (although tweets in general tend to be negative) with the non-experts being more optimistic than the experts. These conclusions cannot be nor confirmed, nor disproved in this thesis because it it not clear what exactly does the attitude in the tweets about (e.g., whether it is pessimism that AI will not develop fast enough or whether these are worries of potential misuse).

In general, I have observed two seemingly contradictory trends caused probably by marketing effort and generally growing interest of technological companies in deep learning.

On one hand, results of deep learning are commonly labeled as artificial intelligence with an unspoken connotation that is is the ultimate goal and the holy grail of computer science. Putting the technology into a sci-fi discourse leads to worries belonging to that discourse. On the other hand, the worries are being calmed down by another trend which is a tendency to report on deep learning as if were a standard innovation seen many times in human history. This may lead to an argumentation fallacy that obfuscate the potential problems brought by the technology. The discourse suggest what should users worry about (AI getting out of control) and at the same time they are affirmed it just a technology like any other. Potential privacy violation or criminal misuse of the technology are excluded from public discussion because they are drown out by seemingly more serious issues.

These trends also create new challenges for scientific community. The publication culture in various fields of computer science affected by deep learning has significantly changed during the studied period. New results are published immediately on arXiv without undergoing a peer review (Acharya et al., 2014). Relevancy of the publication for the community is thus verified mostly via the empirical improvement brought by presented methods. This is indeed convenient for technological companies which can very quickly re-implement the most promising innovations and deploy them in production. Researchers do not want to lose their status as those who drive the field forward and take part in this accelerating publication endeavor without any need to more deeply reflect the achieved results. However, if the scientific community wants to retain its scientific standards, it cannot succumb to this tempting trend.

Despite previously mentioned slightly disturbing conclusions about diverting public attention from potential risks of the technology and challenges that the development poses to the scientific community, there is one more important thing that should not be missed. The interest of technological companies and demand for more and more positive results that get become immediately public contributes to the fact that the field remains remarkably open. Most of the innovations are based on publicly available results which are discussed on the Internet by both experts and interested members of public. This is probably the first time in the modern history when a crucial innovation emerged under full public control without keeping the principles in secret as long as possible or protecting them by patents.

Bibliography

Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., … Zheng, X. (2016). Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *CoRR*, *abs/1603.04467*. Retrieved from `http://arxiv.org/abs/1603.04467`

Acharya, A., Verstak, A., Suzuki, H., Henderson, S., Iakhiaev, M., Lin, C. C., & Shetty, N. (2014). Rise of the rest: The growing impact of non-elite journals. *CoRR*, *abs/1410.2217*. Retrieved from `http://arxiv.org/abs/1410.2217`

Amodei, D., Ananthanarayanan, S., Anubhai, R., Bai, J., Battenberg, E., Case, C., … others (2016). Deep speech 2: End-to-end speech recognition in english and mandarin. In *33rd international conference on machine learning (icml 2016)* (pp. 173–182). New York, NY, USA.

Bahdanau, D., Cho, K., & Bengio, Y. (2014). Neural machine translation by jointly learning to align and translate. *CoRR*, *abs/1409.0473*. Retrieved from `http://arxiv.org/abs/1410.0473`

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation learning: A review and new perspectives. *IEEE Transactions on pattern analysis and machine intelligence*, *35*(8), 1798–1828.

Berger, P., & Luckmann, T. (1966). *The social construction of reality: A treatise in the sociology of knowledge*. Doubleday.

Bergstra, J., Breuleux, O., Bastien, F., Lamblin, P., Pascanu, R., Desjardins, G., … Bengio, Y. (2010, June). Theano: A cpu and gpu math compiler in python. In *Proceedings of the 9th python in science conference* (pp. 1–7). Austing, TX, USA: NumFOCUS Foundation.

Bijker, W. E. (1997). *Of bicycles, bakelites, and bulbs: Toward a theory of sociotechnical change*. MIT Press.

Bijker, W. E. (2006). Why and how technology matters. In R. E. Goodin & C. Tilly (Eds.), *Oxford handbook of contextual political analysis* (pp. 681–706). Oxford, United Kingdom: Oxford Univesity Press.

Bijker, W. E. (2010). How is technology made?–That is the question! *Cambridge Journal of Economics*, *34*(1), 63-76.

Bishop, C. (2007). *Pattern recognition and machine learning (information science and statistics)*. Springer, New York.

Blei, D. M. (2012, April). Probabilistic topic models. *Commun. ACM*, *55*(4), 77–84. Retrieved from http://doi.acm.org/10.1145/2133806.2133826 doi: 10.1145/2133806.2133826

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993–1022.

Carpentier, N. (2012). The concept of participation. if they have access and interact, do they really participate? *Revista Fronteiras*, *14*(2), 13–36.

Curtis, S. (2016, July 1). *Google photos labels black people as 'gorillas'.* Jersey, United Kingdom: Telegraph Media Group Limited. Retrieved from http://www.telegraph.co.uk/technology/google/11710136/Google-Photos-assigns-gorilla-tag-to-photos-of-black-people.html (Accessed: 17/07/2016)

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., & Fei-Fei, L. (2009, June). ImageNet: A large-scale hierarchical image database. In *Computer vision and pattern recognition, 2009. cvpr 2009. ieee conference on* (pp. 248–255). Miami, FL, USA: IEEE Computer Society.

Fairclough, N. (1995). *Critical discourse analysis: Papers in the critical study of language*. Longman.

Fast, E., & Horvitz, E. (2017). Long-term trends in the public perception of artificial intelligence. In *Proceedings of the thirty-first aaai conference on artificial intelligence (aaai-17)* (pp. 963–969). San Franciso, CA, USA: AAAI.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267–285). Springer.

Gibbs, S. (2014, January 27). *Google buys UK artificial intelligence startup Deepmind for £400m.* London, United Kingdom: Guardian News and Media Limited. Retrieved from https://www.theguardian.com/technology/2014/jan/27/google-acquires-uk-artificial-intelligence-startup-deepmind (Accessed: 29/7/2016)

Given, L. M. (2008). *The SAGE encyclopedia of qualitative research methods*. SAGE Publications.

Glorot, X., Bordes, A., & Bengio, Y. (2011, April). Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on* *artificial intelligence and statistics (aistats-11)* (Vol. 15, p. 315-323). Fort Lauderdale, FL, USA: JMLR.org. Retrieved from http://www.jmlr.org/proceedings/papers/v15/glorot11a/glorot11a.pdf

Gong, Y., Lazebnik, S., Gordo, A., & Perronnin, F. (2013). Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(12), 2916–2929.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. Cambridge, MA, USA: MIT press.

Goodman, L. A. (1961). Snowball sampling. *The annals of mathematical statistics*, *32*(1), 148–170.

Graves, A., Wayne, G., & Danihelka, I. (2014). Neural turing machines. *CoRR*, *abs/1410.5401*. Retrieved from http://arxiv.org/abs/1410.5401

Gupta, S. (2014, March 25). *What's Machine Learning? Thanks to GPU Accelerators, You're Already Soaking In It* . Santa Clara, CA, USA: Nvidia Corporation. Retrieved from https://blogs.nvidia.com/blog/2014/03/25/machine-learning/ (Accessed: 17/07/2016)

Halliday, J. (2011). *London riots: How blackberry messenger played a key role* (Vol. 8). London, United Kingdom: Guardian News and Media Limited. (Accessed: 29/7/2016)

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, *abs/1512.03385*. Retrieved from http://arxiv.org/abs/1512.03385

Helft, M. (2016, June 15). *Google's bold move to reinvent every device on the planet.* San Francisco, CA, USA: Forbes Media LLC. Retrieved from http://www.forbes.com/sites/miguelhelft/2016/06/15/google-is-about-to-change-everything-again (Accessed 31/7/2016)

Hern, A. (2016, September 28). *'partnership on AI' formed by Google, Facebook, Amazon, IBM and Microsoft.* London, United Kingdom: Guardian News and Media Limited. Retrieved from https://www.theguardian.com/technology/2016/sep/28/google-facebook-amazon-ibm-microsoft-partnership-on-ai-tech-firms (Accessed: 12/11/2016)

Hernandez, D. (2014a, 5 16). *Man behind the 'Google Brain' joins Chinese search giant Baidu.* San Franciso, CA, USA: Condé Nast Publications. Retrieved from https://www.wired.com/2014/05/andrew-ng-baidu/ (Accessed 10/7/2017)

Hernandez, D. (2014b, January 16). *Meet the man Google hired to make AI a reality.* San Franciso, CA, USA: Condé Nast Publications. Retrieved from http://www.wired.com/2014/01/geoffrey-hinton-deep-learning (Accessed 31/7/2016)

High, P. (2016, June 20). *Deep learning pioneer Geoff Hinton helps shape Google's drive to put AI everywhere.* San Francisco, CA, USA: Forbes Media LLC. Retrieved from http://www.forbes.com/sites/peterhigh/2016/06/20/deep-learning-pioneer-geoff-hinton-helps-shape-googles-drive-to-put-ai-everywhere (Accessed 31/7/2016)

Hinton, G. E., Deng, L., Yu, D., Dahl, G., rahman Mohamed, A., Jaitly, N., … Kingsbury, B. (2012, November). Deep neural networks for acoustic modeling in speech recognition. *IEEE Signal Processing Magazine*, *29*(6), 82–97.

Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006, July). A fast learning algorithm for deep belief nets. *Neural Computation*, *18*(7), 1527–1554.

Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2012). Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*, *abs/1207.0580*. Retrieved from http://arxiv.org/abs/1207.0580

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735–1780.

Ioffe, S., & Szegedy, C. (2015, July). Batch normalization: Accelerating deep network training by reducing internal covariate shift. In F. R. Bach & D. M. Blei (Eds.), *Proceedings of the 32nd international conference on machine learning, ICML 2015* (Vol. 37, p. 448-456). Lille, France: JMLR.org. Retrieved from http://proceedings.mlr.press/v37/ioffe15.pdf

Jouppi, N. (2016, May 18). *Google supercharges machine learning tasks with TPU custom chip.* Mountain View, CA, USA: Google, Inc. Retrieved from https://cloudplatform.googleblog.com/2016/05/Google-supercharges-machine-learning-tasks-with-custom-chip.html (Accessed: 17/07/2016)

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012, May). Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems* (pp. 1097–1105). Red Hook, NY, USA: Curran Associates, Inc. Retrieved from http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf

Kuhn, T. S. (1970). *The structure of scientific revolutions*. Chicago, MA, USA: University of Chicago Press.

Kvasz, L. (2008). *Patterns of change: Linguistic innovations in the development of classical mathematics*. Birkhäuser Basel.

Latour, B. (1992). Where are the missing masses, sociology of a few mundane artefacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology / building society: Studies in sociotechnical change* (p. 225-258). Cambridge/MA: The MIT Press.

LeCun, Y. (2016, March 24). *Slides on deep learning.* Geneva, Switzerland: CERN. Retrieved from `https://indico.cern.ch/event/510372/`

LeCun, Y., Bengio, Y., & Hinton, G. E. (2015). Deep learning. *Nature*, *521*(7553), 436–444.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324.

LeCun, Y., Bottou, L., Orr, G. B., & Müller, K.-R. (1998). Efficient backprop. In *Neural networks: Tricks of the trade* (pp. 9–48). Springer.

Manikonda, L., & Kambhampati, S. (2017). Tweeting AI: perceptions of lay vs expert twitterati. *CoRR*, *abs/1709.09534*. Retrieved from `http://arxiv.org/abs/1709.09534`

McLuhan, M. (1964). *Understanding media: the extensions of man*. McGraw-Hill.

Metz, C. (2013, 12 9). *Facebook taps 'deep learning' giant for new AI lab.* San Franciso, CA, USA: Condé Nast Publications. Retrieved from `https://www.wired.com/2013/12/facebook-yann-lecun/` (Accessed 10/7/2017)

Metz, C. (2015, December 12). *Elon Musk's billion-dollar AI plan is about far more than saving the world.* San Franciso, CA, USA: Condé Nast Publications. Retrieved from `https://www.wired.com/2015/12/elon-musks-billion-dollar-ai-plan-is-about-far-more-than-saving-the-world/`

Metz, C. (2016, 4 5). *Facebook's AI is now automatically writing photo captions.* San Franciso, CA, USA: Condé Nast Publications. Retrieved from `https://www.wired.com/2016/04/facebook-using-ai-write-photo-captions-blind-users/` (Accessed 9/7/2017)

Mikolov, T., Yih, W.-t., & Zweig, G. (2013, June). Linguistic regularities in continuous space word representations. In *Proceedings of the 2013 conference of the north american chapter of the association for computational linguistics: Human language technologies* (pp. 746–751). Atlanta, Georgia: Association for Computational Linguistics. Retrieved from http://www.aclweb.org/anthology/N13-1090

Moody, C. E. (2016). Mixing dirichlet topic models and word embeddings to make lda2vec. *CoRR*, *abs/1605.02019*. Retrieved from http://arxiv.org/abs/1605.02019

O'Regan, G. (2016). *Introduction to the history of computing: A computing history primer*. Springer International Publishing.

Řehůřek, R., & Sojka, P. (2010, May 22). Software framework for topic modelling with large corpora. In *Proceedings of the LREC 2010 workshop on new challenges for NLP frameworks* (pp. 45–50). Valletta, Malta: ELRA. Retrieved from http://is.muni.cz/publication/884893/en

Rosenblatt, F. (1957). *The perceptron, a perceiving and recognizing automaton project para* (Tech. Rep.).

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, *65*(6), 386–408.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1988). Neurocomputing: Foundations of research. In J. A. Anderson & E. Rosenfeld (Eds.), (pp. 696–699). Cambridge, MA, USA: MIT Press. Retrieved from http://dl.acm.org/citation.cfm?id=65669.104451

Russell, S. J., & Norvig, P. (2002). *Artificial intelligence: A modern approach (2nd edition)*. Prentice Hall.

Samuel, A. L. (1959). Some studies in machine learning using the game of checkers. *IBM Journal of research and development*, *3*(3), 210–229.

Sang-Hun, C. (2016, 3 15). *Google's computer program beats Lee Se-dol in Go tournament.* New York City, NY, USA: The New York Times Company. Retrieved from https://www.nytimes.com/2016/03/16/world/asia/korea-alphago-vs-lee-sedol-go.html (Accessed 9/7/2017)

Schmidhuber, J. (2015, June). *Critique of paper by "Deep Learning Conspiracy" (Nature 521 p. 436).* Lugano, Switzerland: University of Applied Sciences and Arts of Italian Switzerland. Retrieved from http://people.idsia.ch/~juergen/deep-learning-conspiracy.html

Searle, J. R. (1980, September). Minds, brains, and programs. *Behavioral and Brain Sciences*, *3*(3), 417–424.

Sennrich, R., Haddow, B., & Birch, A. (2016, August). Neural machine translation of rare words with subword units. In *Proceedings of the 54th annual meeting of the association for computational linguistics (volume 1: Long papers)* (pp. 1715–1725). Berlin, Germany: Association for Computational Linguistics. Retrieved from `http://www.aclweb.org/anthology/P16-1162`

Sgall, P., Hajicová, E., Panevová, J., & Mey, J. (1986). *The meaning of the sentence in its semantic and pragmatic aspects*. Dordrecht, The Netherlands: Springer Netherlands.

Silver, D., & Hassabis, D. (2016, 1 27). *AlphaGo: Mastering the ancient game of go with machine learning.* Mountain View, CA, USA: Google, Inc. Retrieved from `https://research.googleblog.com/2016/01/alphago-mastering-ancient-game-of-go.html` (Accessed: 29/7/2016)

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., van den Driessche, G., … Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, *529*(7587), 484–503.

Simonite, T. (2017, August 31). *Google and Microsoft can use ai to extract many more ad dollars from our clicks.* San Franciso, CA, USA: Condé Nast Publications. Retrieved from `https://www.wired.com/story/big-tech-can-use-ai-to-extract-many-more-ad-dollars-from-our-clicks` (Accessed 30/9/2017)

Srivastava, R. K., Greff, K., & Schmidhuber, J. (2015). Highway networks. *CoRR*, *abs/1505.00387*. Retrieved from `http://arxiv.org/abs/1505.00387`

Steele, B. (2012, September). *Library-managed 'arXiv' spreads scientific advances rapidly and worldwide.* Ithaca, NY, USA: Cornell University. Retrieved from `http://ezramagazine.cornell.edu/FALL12/CoverStorySidebar2.html` (Accessed 5/10/2016)

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. *CoRR*, *abs/1409.3215*. Retrieved from `http://arxiv.org/abs/1409.3215`

Tokui, S., Oono, K., Hido, S., & Clayton, J. (2015, December). *Chainer: a next-generation open source framework for deep learning.* Retrieved from `http://learningsys.org/papers/LearningSys_2015_paper_33.pdf`

Turing, A. M. (1950). *Computing machinery and intelligence* (Vol. 59). Oxford, United Kingdom: Oxford University Press.

Vaughan, D. (1997). *The challenger launch decision: Risky technology, culture, and deviance at NASA*. University of Chicago Press.

Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. (2015, June). Show and tell: A neural image caption generator. In *IEEE conference on computer vision and* *pattern recognition, CVPR 2015* (pp. 3156–3164). Boston, MA, USA: IEEE. Retrieved from http://dx.doi.org/10.1109/CVPR.2015.7298935

Wu, Y., Schuster, M., Chen, Z., Le, Q. V., Norouzi, M., Macherey, W., … Dean, J. (2016). Google's neural machine translation system: Bridging the gap between human and machine translation. *CoRR*, *abs/1609.08144*. Retrieved from http://arxiv.org/abs/1609.08144

Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A. C., Salakhutdinov, R., … Bengio, Y. (2015, July). Show, attend and tell: Neural image caption generation with visual attention. In *Proceedings of the 32nd international conference on machine learning, ICML 2015* (pp. 2048–2057). Lille, France: JMLR.org. Retrieved from http://jmlr.org/proceedings/papers/v37/xuc15.html